CDF///CDFR/9292
Version 2.00
November 12, 2008

Efficiency Measurement of the RomaNN b-Tagger

Justin Keung[1], Chris Neu, Evelyn Thomson
University of Pennsylvania

Abstract

This note describes a measurement of the efficiency for identifying b-jets in data using the Roma Neural Net (RomaNN) b-tagger. This tool is used to make a per-jet b-tag decision (binary mode RomaNN) by placing a cut on the RomaNN output value. A sample of b-enriched jets is collected in the data by looking for jets containing a semi-leptonic decay; the transverse momentum of the lepton relative to the jet vector is used to discriminate between b and non-b-jets. The fractions of these types of jets in "tagged" and "not-tagged" samples are used to calculate the tag efficiency for b-jets. This data efficiency for tagging b-jets is compared to the efficiency for tagging b-jets in Monte Carlo (MC) simulated events, and the resulting mismatch in data and simulation efficiencies is expressed as a scale factor, which can then be used to obtain agreement between the simulation and the data. The operating point optimized for the WH analysis, RomaNN3out>0.00, has a scale factor of $(0.8821 - 0.006541 \times (JetE_T(inGeV) - 45)) \pm 0.061(stat{+}JetE_T\, param) \pm 0.028(syst)$ for all muon jets, and for all other jets a scale factor of
$0.793 \times \frac{0.8771-0.02285\times(nZvertex-1.8)}{0.8771} \pm 0.018(stat) \pm 0.030(nZvertex) \pm 0.080(syst)$.

Contents

[1]keungj@fnal.gov

1 Motivation: The RomaNN b-Tagger

At CDF it is often necessary to identify jets that come from b quarks, so-called b-tagging. The identification of b-jets is an essential component for measurements in the top quark sector and searches for a low mass Higgs boson and other new phenomena. The signatures of these interesting signal processes all contain b jets; the ability to discriminate b jets from the overwhelming inclusive jet background helps increase the purity of the selected event sample. For a signature like $WH \to l\nu b\bar{b}$, the most useful sample has two b-tagged jets: the number of such signal events is proportional to the square of the efficiency for tagging b-jets. The motivation for this study is to calibrate the efficiency of a new b-tag algorithm [4].

The Roma neural net b-tagger (RomaNN) is a new b-tag algorithm [4] that has demonstrated improved performance over SecVtx [8] and JetProb [9], the standard CDF b-tagging algorithms. Fig. 1 shows the efficiency for tagging light-jets as a function of the efficiency for tagging b-jets for the Tight/Loose SecVtx and RomaNN taggers. If one draws a horizontal line, then an indicator of the performance of a b-tagger is how far to the right the curve intersects it. The further to the right the intersection indicates better separation power between b-jets and light-jets, since for the same efficiency for tagging light-jets it has a higher efficiency for tagging b-jets. For example, in a $p\bar{p} \rightarrow q\bar{q}$ di-jet MC, the Tight SecVtx has an efficiency for tagging light-jets of 0.017, and achieving an efficiency for tagging b-jets of 0.4; whereas the RomaNN tagger achieves an efficiency for tagging b-jets of 0.53 at the same efficiency for tagging light-jets of 0.017.

This note describes the calibration of the efficiency for tagging b-jets in data. This is necessary since simulations may not adequately model the detector and result in an over-estimate or under-estimate of the efficiency for tagging b-jets. Previous studies for SecVtx used semileptonic B hadron decays from a data sample triggered by a muon, and found a data/MC correction factor of 0.932±0.022. In this study, we use a data sample triggered by an electron to avoid any bias due to the RomaNN algorithm using muons as a discriminating observable between heavy and light flavor hadrons.

RomaNN is different from the standard tagging algorithms in that it is a multivariate tool, incorporating information from several sources simultaneously. It incorporates information from SecVtx, JetProb, soft muon from semileptonic decays, as well as other variables in its per-jet tagging decision, to achieve a net gain of 30% in the efficiency for tagging b-jets [5] in simulated Monte Carlo (MC) events. RomaNN provides a per-jet output value in the range from -1 to +1; a value near +1 indicates that the jet is consistent with coming from a b and values away from +1 indicate the jet is more consistent with originating from some other flavor. Each analysis can choose its own cut in the RomaNN output value (binary mode RomaNN), giving an opportunity to customize the level of b purity.

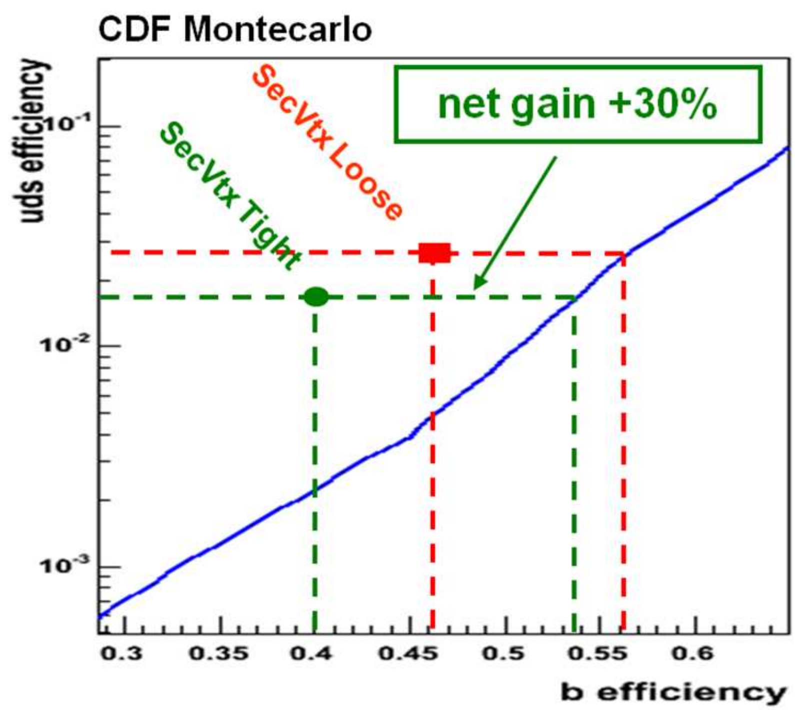


Figure 1: Efficiency for tagging light-jets as a function of efficiency for tagging b-jets for the Tight/Loose SecVtx and RomaNN taggers [5].

The RomaNN consists of three Neural Networks, for vertex identification (VerticesNN), track identification (TracksNN), and flavor identification. The Neural Networks are based on the commercial NeuroBayes package [1]. Fig. 2 gives an overview of the information flow within RomaNN.

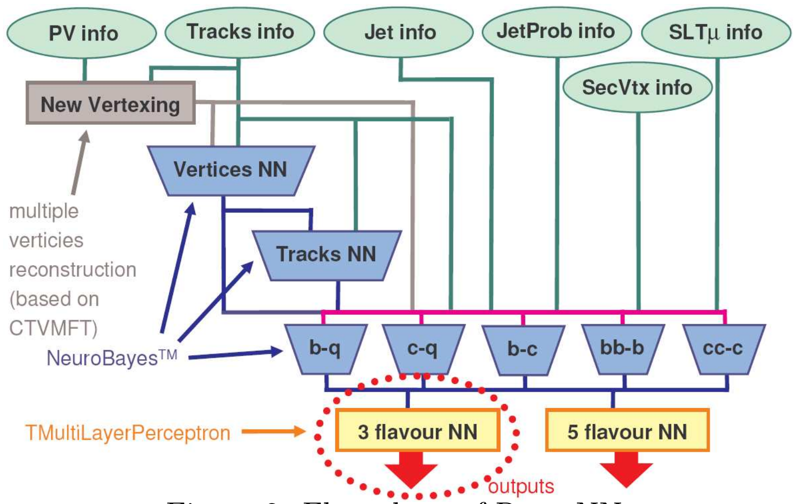


Figure 2: Flow chart of RomaNN.

A first Neural Network VerticesNN makes use of a vertexing algorithm based on the CTVMFT package, which associates the collection of tracks inside a jet to a set of vertices, and is used to distinguish the vertices produced by the decay of heavy flavor from light flavor hadrons. For the tracks not used by VerticesNN, a second Neural Network TracksNN is used to distinguish between tracks produced by heavy hadron decay and prompt tracks. The information from these two Neural Networks is combined with information from existing CDF b-tag tools SecVtx, JetProb, and SLTμ to form a third Neural Network, giving two output numbers. One of the output number is "RomaNN3out". Fig. 3 is the output of RomaNN for MC jets passing W+2jets selection, showing that b-jets have output higher values than charm or light-jets.

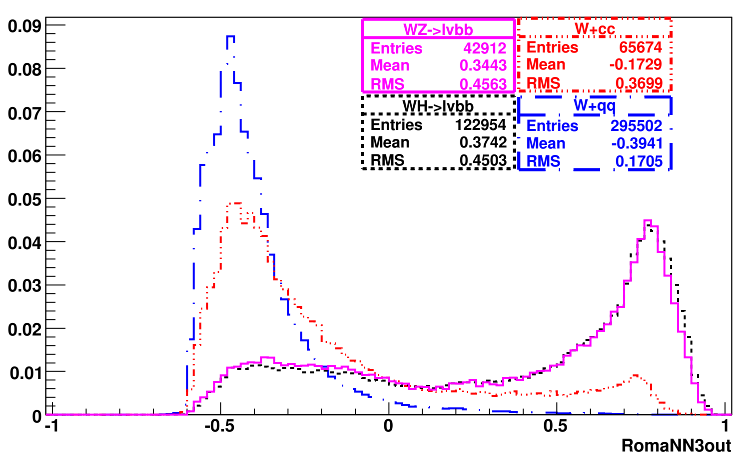


Figure 3: RomaNN3out of MC jets passing W+2jets selection, normalized to unit area: Signal B jets from WH(120GeV) and WZ, C jets from W+cc and LF jets from W+qq. Flavour is matched to OBSP hadron within deltaR of 0.4.

The strength of the RomaNN b-tag algorithm lies in its ability to use all useful information. The input variables used in vertex, track, and flavor indentification are listed in tables 1, 2, and 3 respectively. These variables all have significance $> 5\sigma$.

Rank	*Variable*
1	Vertex Pseudo $c\tau$ (secondary and primary vertex separation)
2	Significance of the d_0 of the 2nd Most Displaced Track in Vertex
3	Angle Between Vertex Momentum Vector and Vertex Displacement Vector
4	L_{xyz} Significance
5	Invariant Mass of Vertex

Table 1: Input variables to the vertex NN, ranked by NeuroBayes.

Rank	*Variable*
1	track d_0 significance
2	D/L
3	vertex NN
4	α
5	D

Table 2: Input variables to the track NN, ranked by NeuroBayes. Variables defined in Fig. 4

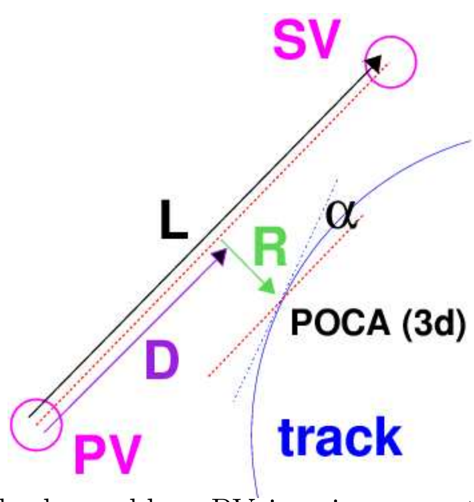


Figure 4: Definition of track observables. PV is primary vertex. SV is secondary vertex. POCA is point of closest approach in 3d of the track to the line segment connectinging PV to SV.

2 Efficiency Measurement: Lepton p_T^{rel} Technique

One of the most important performance parameters of the identification of b-jets is the efficiency at which they are correctly identified. With this efficiency measurement one can estimate in simulated Monte Carlo (MC) events the yield of signal and background as one would expect in an analysis. However, this efficiency needs to be measured in the data, since there is no guarantee that the efficiency measured in simulated events is accurate.

Rank	*Variable*
1	JetProbability (when calculated with at least 2 tracks)
2	Pseudo $c\tau$ of Best Vertex
3	LooseSecVtx Tag
4	Number of Muons Identified By SLTμ
5	$\Sigma_{selected\ tracks} p_T / \Sigma_{all\ jet\ tracks} p_T$ (scalar sums)
6	L_{xyz} Significance of Best Vertex
7	Invariant Mass of Best Vertex
8	Invariant Mass of Selected Tracks (both Vertexed and Unvertexed)
9	Number Of Selected Tracks
10	L_{xyz} of Best Vertex
11	Number Of Good Tracks In Jet
12	$\Sigma_{all\ jet\ tracks} p_T$ (scalar sums)
13	L_{xyz} of Second Best Vertex
14	p_T of (Highest p_T) Muon With Respect To Jet Axis
15	Mass of SecVtx Vertex
16	Jet E_T

Table 3: Input variables to the flavor NNs, ranked by NeuroBayes.

The approach widely used at CDF is to measure the efficiency for tagging b jets (Eqn. 4) in the data and in the simulation, and encode any mismatch in a data/MC correction factor. We follow this approach, and describe in this note a measurement of the data-to-MC scale factor for the binary mode RomaNN.

In MC, the efficiency for tagging b-jets measurement can be calculated simply by counting the b content of a sample from the OBSP bank, before and after applying the b-tag.

In data, the technique we use relies on the fact that the so-called lepton jets (data jets resulting from heavy flavor decay $B \to \mu X$ or $B \to eX$, where B represents a B hadron) will on average contain a lepton with a higher transverse momentum relative to the jet vector compared to the decay of a generic hadron, since B hadrons have a higher invariant mass than generic hadrons. The lepton transverse momentum relative to the jet vector (p_T^{rel}) is defined in Eqn. 1.

$$lepton\ p_T^{rel} = \left|\overrightarrow{P}_{lepton}\right| \sqrt{1 - \left(\frac{\overrightarrow{P}_{lepton} \cdot \overrightarrow{P}_{jet}}{\left|\overrightarrow{P}_{lepton}\right| \left|\overrightarrow{P}_{jet}\right|}\right)^2} \tag{1}$$

where $\overrightarrow{P}_{lepton}$ is the momentum of the lepton and $\overrightarrow{P}_{jet}$ is the momentum of the lepton-jet.

We use both the electron and the muon as the lepton. This will be explained in the following section. The momentum of the electron is obtained from a combination of tracking and calorimetry, processed through the "TopAlgorithms::correctElectron" algorithm. The momentum of the muon is obtained from tracking. The energy and momentum of the electron-jet is obtained from a combination of tracking and calorimetry, processed through the "JetCorr" algorithm. The energy and momentum of the muon-jet, however, must be corrected for the muon escaping the calorimeter with most of its momentum according to Eqn. 2 and 3. This was not necessary for electron, since electrons do not escape.

$$\overrightarrow{P}_{corr} = \overrightarrow{P}_{jet} + \left(1 - \frac{2\,\frac{GeV}{c}}{|p_\mu|}\right) \overrightarrow{P}_\mu \tag{2}$$

$$E_{T,corr} = E_{T,jet}\left(\frac{E_{jet} - 2\ GeV}{E_{jet}}\right) + p_{T,\mu} \tag{3}$$

The b content of a sample can be determined by using the discriminating shape of the p_T^{rel} between b and non-b jets. Splitting the data sample into two subsamples, "tagged" and "not-tagged", we can fit their p_T^{rel} distribution to find the b content of each. This allows us to calculate the efficiency for tagging b-jets in data from the yields of b-jets in each subsample (Eqn. 4).

$$\epsilon_b = \frac{N^b_{Tag}}{N^b_{total}} = \frac{N^b_{Tag}}{N^b_{Tag} + N^b_{NoTag}} \tag{4}$$

where N^b_{Tag} is the number of b-jets that are tagged (RomaNN3out>RomaNNcut); N^b_{NoTag} is the number of b-jets that are not tagged (RomaNN3out≤RomaNNcut, including RomaNN not-taggable jets); N^b_{total} is the total number of b-jets. Then we can obtain the data-to-MC scale factor by dividing the efficiency for tagging b-jets in data by the efficiency for tagging b-jets in MC (Eqn. 5).

$$SF = \frac{\epsilon_{data}}{\epsilon_{MC}} \tag{5}$$

2.1 Electron vs Muon

The p_T^{rel} technique has been used in CDF since 2005 [3] to measure the efficiency for tagging b-jets, with the jets having a muon. These jets can represent the generic candidate jets for the SecVtx b-tagger because the tagger does not explicitly use the muon information. However, because the RomaNN uses explicitly the number of muons as an input, jets with a muon are no longer good handles to measure the efficiency for tagging b-jets in general.

For this reason, we have developed the technique further, to use jets having an electron to measure the efficiency for tagging b-jets. The RomaNN3out distributions from b-hadronic (no muon and no electron) jets, b-jets containing an electron, and b-jets containing a muon are shown in Fig. 5. It is clear that generic jets and jets with electrons have similar distributions, while jets with a muon have a distribution shifted towards positive (b-like) values. Because of this, we measure the scale factor of two sub-samples of jets: muon-jets, and generic jets as represented by electron-jets.

But how do we justify representing generic jets with electron jets? In a di-jet sample where events have "back to back" jets, with both jets have the same kinematic requirements, we require one jet to have an electron. We require that both jets have RomaNN3out>0.0, such that this sample is enriched in b-content. The jets with electrons are referred to as the electron-jets, the other is referred to as the away-jets. The away-jets can represent generic jets because they do not have any lepton requirements. In Fig. 6 we compare the NN3out distributions of the electron-jets and away-jets. Fig. 6 a and b shows visually the level of agreement between the data and MC NN3out distributions. Fig. 6 c and d shows that the difference between the electron-jets and the away-jets exist in both data and MC.

Fig. 7 shows the integral of the NN3out distribution of data/MC and electron/away-jets starting from RomaNNcut=0.0, which is the number of jets tagged.

Fig. 8 shows the quotient of the number of away/electron-jets tagged in data and number of away/electron-jets tagged in MC. To be sure these samples are not purely b-jets, but

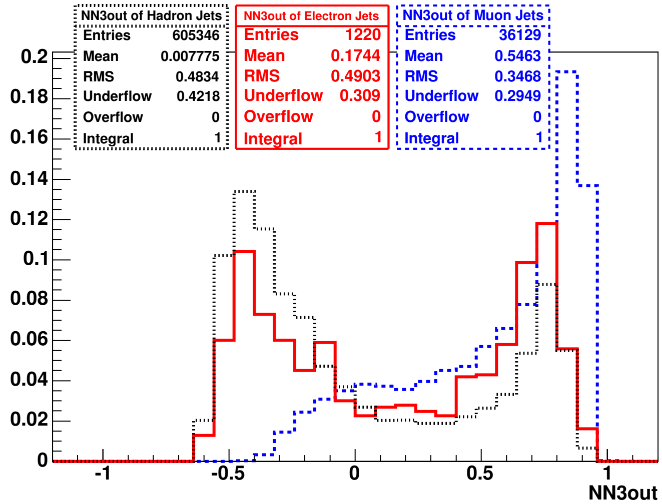


Figure 5: RomaNN3out distribution comparison between generic, electron, and muon, from WZ Monte Carlo with jet $|\eta| < 1.2$ and $E_T > 9$ GeV.

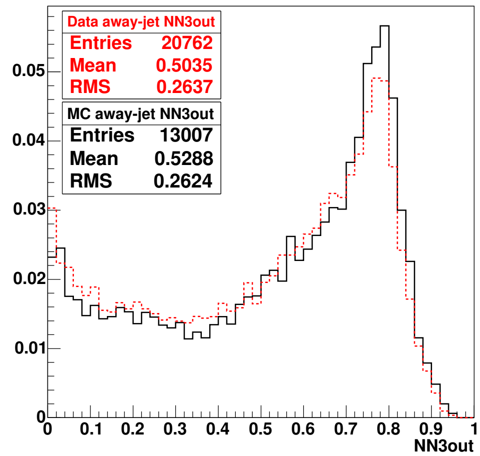
these quotient are scale factors between the electron-jet sample and the away-jet sample, and equally contaminated. Their difference is shown in Fig. 9, showing that the difference in equally contaminated scale factors of the away-jets and the electron-jets is small, justifying our extrapolation of the generic-jets from electron-jets.

3 Event Selection

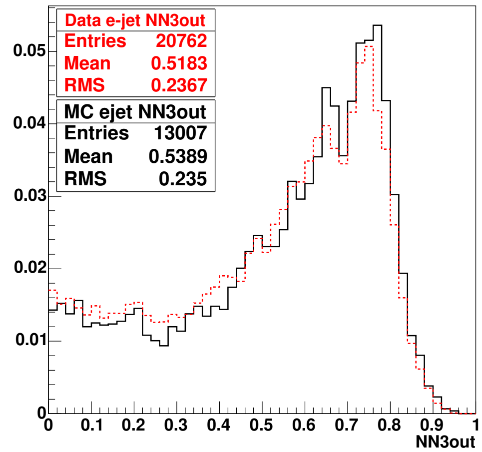
To measure efficiency for tagging b-jets, it is necessary to get a sample of jets with enriched b content. To do that, we select events with "back to back" ($|\phi_{probe} - \phi_{away}| > 2.0$ radians) jets (Fig. 10), having a probe-jet and an away-jet, requiring that the away-jet be LooseSecVtx tagged. We require that the probe-jet contain a lepton within a cone of 0.4 in η–ϕ space of the center of the jet. Should there be more than one lepton inside the candidate probe-jet, the lepton with the highest E_T is selected. The selection criteria for electron-jets, muon-jets, and away-jet are listed in tables 4, 5 and 6.

3.1 Samples

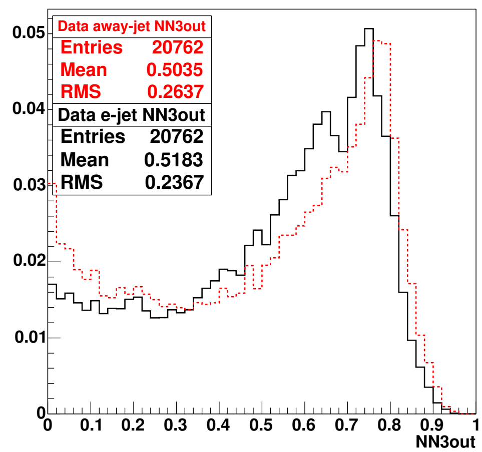
The MC samples correspond to an integrated luminosity of $1.5fb^{-1}$. The generating process is Pythia di-jet $P_T > 20$ GeV, $|\eta| < 2.0$, filtered for the electron-jet sample to have an electron with $P_T > 8$ GeV and $|\eta| < 1.2$, and filtered for the muon-jet sample to have a muon with $P_T > 9$ GeV and $|\eta| < 0.6$. The data samples are taken from the 8 GeV Electron trigger path and the 8 GeV Muon trigger path, with an integrated luminosity of $3fb^{-1}$. See table 7 for definition. Events in the good runs list version 24 are selected (up to period 18). Good quality silicon, muon and EM calorimeter runs are required for data; good silicon and EM calorimeter runs are required in MC (the muon requirement not being modeled in Monte Carlo). Level 5 Jet energy correction from the JetCorr17 package are applied to the jets. The probe/away-jet pair yields from the above samples are listed in tables 10, 11, 8 and 9.



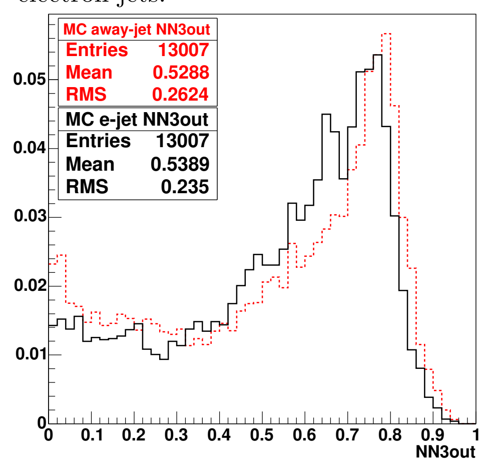
(a) Comparison of the NN3out distribution of data away-jets and MC away-jets.



(b) Comparison of the NN3out distribution of data electron-jets and MC electron-jets.

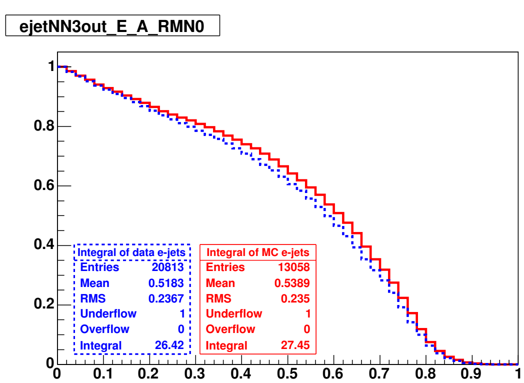


(c) Comparison of the NN3out distribution of data electron-jets and data away-jets.

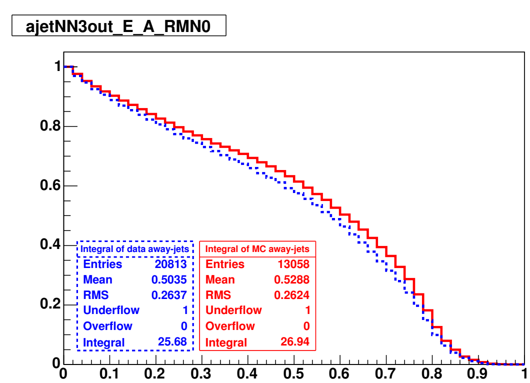


(d) Comparison of the NN3out distribution of MC electron-jets and MC away-jets.

Figure 6: Comparison of the NN3out distribution of data/MC electron/away-jets.

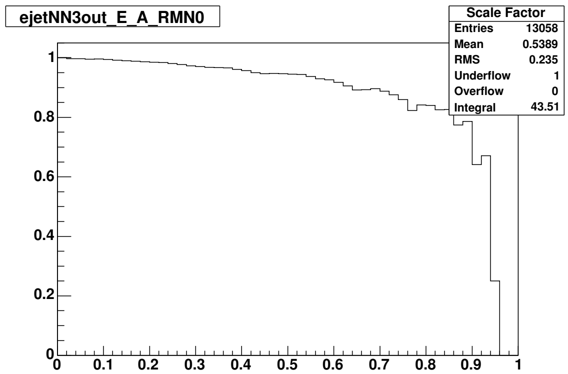


(a) Integral of the NN3out distribution of data and MC electron-jets.

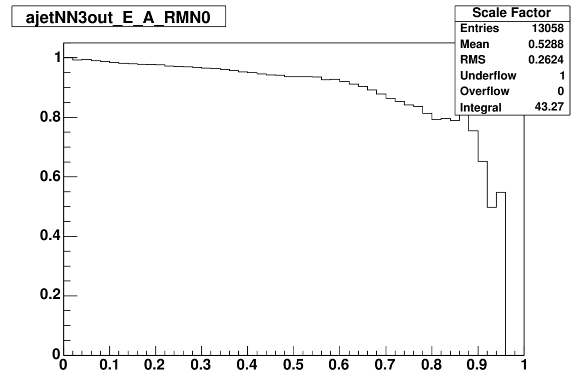


(b) Integral of the NN3out distribution of data and MC away-jets.

Figure 7: Integral of the NN3out distribution of data/MC electron/away-jets starting from RomaNNcut=0.0.



(a) Quotient of the integrals between data and MC electron-jets.



(b) Quotient of the integrals between data and MC away-jets.

Figure 8: Quotient of the integrals between data and MC.

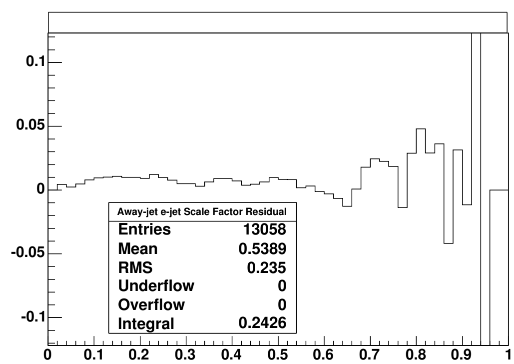


Figure 9: Difference of the away-jets and electron-jets quotient.

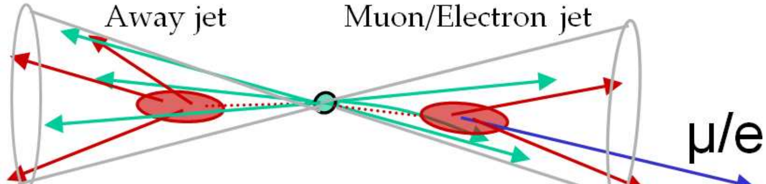


Figure 10: Probe-jet (muon/electron) and away-jet.

	Requirements	
-	raw Jet $E_T > 15$ GeV	
-	Contains An Electron (ΔR < 0.4)	
	-	Electron CollType 1 (defEmObject)
	-	CEM Electron ($\lvert\eta\rvert < 1.2$)
	-	$E_T > 9$ GeV
	-	$P_T > 8$ GeV
	-	0.5 < E/p < 2
	-	Had/em < 0.05
	-	Lshr < 0.2
	-	$\lvert$Signed CES $\Delta x\rvert < 3$ cm
	-	$\lvert$CES $\Delta z\rvert < 5$ cm
	-	Strip $\chi^2 < 10$
	-	z_0 Within 5 cm of Primary Vertex
	-	Isolation > 0.1
	-	Fiducial to SVX
	-	Conversion Veto

Table 4: Electron-jet requirements.

	Requirements
-	raw Jet $E_T > 15$ GeV
-	Contains an SLT Muon
-	Muon track $\chi^2 < 2$

Table 5: Muon-jet requirements.

	Requirements
-	Jet Energy Level 5 Corrected > 15 GeV
-	$\lvert\eta\rvert < 1.5$
-	$\lvert\phi_{Away} - \phi_{Probe}\rvert > 2.0$ radians
-	LooseSecVtx Tagged

Table 6: Away-jet requirements.

NTupleSet	$\int \mathcal{L}$	*Description*
blpc data	3 fb^{-1}	8 GeV Electron Data
		(run range 138425-264071)
bmcl data	3 fb^{-1}	8 GeV Muon Data
		(run range 138425-264071)
btopla di-jet MC	1.5 fb^{-1}	Pythia di-jet $P_T > 20$ GeV, $\lvert\eta\rvert < 2.0$
		Electron $P_T > 8$ GeV, $\lvert\eta\rvert < 1.2$
		Muon $P_T > 9$ GeV, $\lvert\eta\rvert < 0.6$

Table 7: Data and Monte Carlo samples used to measure the Scale Factor.

NTupleSet	*# b RomaNN>0.0*	*# b RomaNN≤0.0*	*# c*
btopla di-jet MC	19317 pairs	24945 pairs	88462 pairs

Table 8: Monte Carlo yields for electron-jets.

NTupleSet	*# RomaNN>0.0*	*# RomaNN≤0.0*
blpc data	66864 pairs	180549 pairs

Table 9: Data yields for electron-jets.

NTupleSet	*# b RomaNN>0.0*	*# b RomaNN≤0.0*	*# c*
btopla di-jet MC	35480 pairs	4440 pairs	176230 pairs

Table 10: Monte Carlo yields for muon-jets.

NTupleSet	*# RomaNN>0.0*	*# RomaNN≤0.0*
bmcl data	172613 pairs	101242 pairs

Table 11: Data yields for muon-jets.

4 Lepton p_T^{rel} Procedures

The lepton p_T^{rel} procedures consist of constructing a sample data distribution, a signal template, and background templates, then fitting for the signal fraction in the data sample.

4.1 MC Electron p_T^{rel} Templates Construction

Sources of electrons include semileptonic B and D hadron decays, fakes from light flavor jets, and electrons from photon conversions in the detector material.

The bottom-jet templates (b-templates) are constructed from the electron-jets from btopla MC, requiring OBSP level match to a b-jet, and the away-jet to be positively tagged by Loose SecVtx; having the same b-enrichment event selection as in data. The RomaNN tag information is used to separate the b-jets, which are then used to construct the tagged and not-tagged electron p_T^{rel} b-templates as shown in Fig. 11. The difference in mean is 0.15 GeV, due to the fact that the tagged sample has on average higher energy jets than the not-tagged sample and the electron p_T^{rel} is proportional to the electron p_T and thus the jet energy.

The charm-jet template (c-template) is constructed from the electron-jets from btopla MC, requiring OBSP level matching to c-jets, without any away-jet requirement. The RomaNN tag information is ignored for the charm-jets. The light-flavor-jet template is constructed from the electron-jets from btopqb and btoprb MC (di-jet MC with P_T >18,40 GeV respectively), veto on OBSP level match to b or c-jets, without any away-jet requirement. The RomaNN tag information is ignored for the light-flavor-jets.

The conversion electron template is constructed from data electron-jets taken from the 8 GeV Electron trigger path, requiring the away-jet to be negatively tagged by Loose SecVtx, and the electron inside the electron-jet to have 0 silicon hits registered where there should have been hits. This sample is dominated by photon conversions where the electron originates in the detector and is not from the primary vertex [10]. The RomaNN tag information is

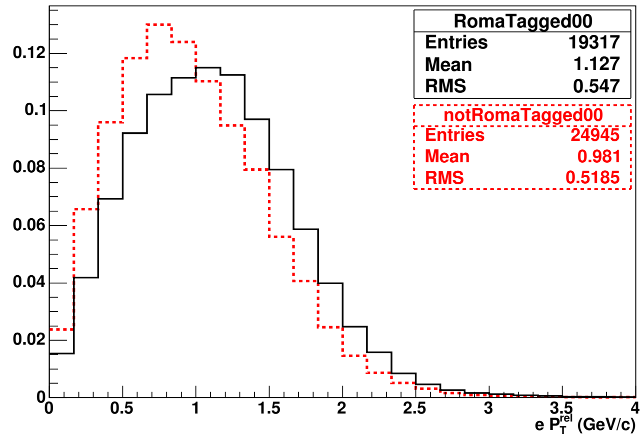


Figure 11: Comparison of the b electron p_T^{rel} templates, normalized to unit area.

ignored for the conversion-jets. The electron p_T^{rel} templates for charm, light-flavor, and conversions are shown in Fig. 12.

The light-flavor template is very similar to the conversion template. Because of this, we can choose either of the two to represent non-b and non-c: we chose the conversion electron template since it has much more statistics than the MC light-flavor-jet template. We will call the conversion electron template our l-template.

In this note, these b,c,l-templates are refered to as the default p_T^{rel} templates.

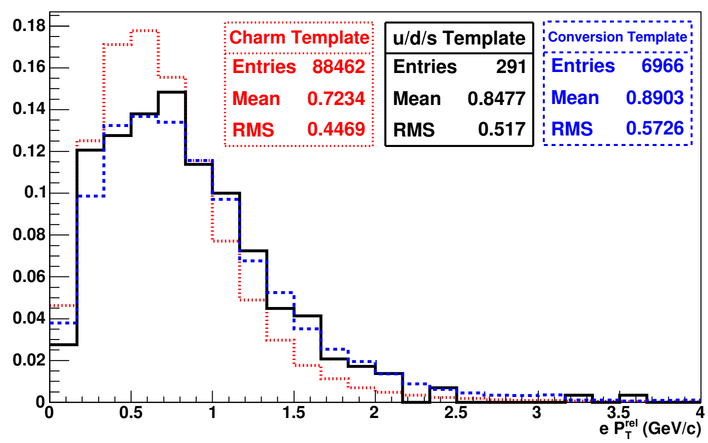


Figure 12: Comparison of the non-b electron p_T^{rel} templates, normalized to unit area.

4.2 MC Muon p_T^{rel} Templates Construction

Sources of muons include semileptonic B and D hadron decays, and fakes from light flavor jets.

The bottom-jet templates (b-templates) are constructed from the muon-jets from btopla MC, requiring OBSP level match to a b-jet, and the away-jet to be positively tagged by Loose SecVtx; having the same b-enrichment event selection as in data. The RomaNN tag information is used to separate the b-jets, which are then used to construct the tagged and not-tagged muon p_T^{rel} b-templates as shown in Fig. 13.

The charm-jet template (c-template) is constructed from the muon-jets from btopla MC, requiring OBSP level matching to c-jets, without any away-jet requirement. The RomaNN

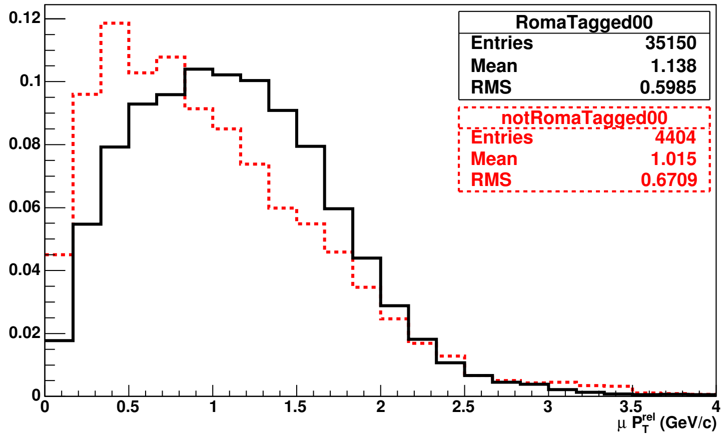


Figure 13: Comparison of the b muon p_T^{rel} templates, normalized to unit area.

tag information is used to separate the c-jets, which are then used to construct the tagged and not-tagged muon p_T^{rel} c-templates. The light-flavor-jet template is constructed from the muon-jets from btopqb and btoprb MC (di-jet MC with P_T >18,40 GeV respectively), veto on OBSP level match to b or c-jets, without any away-jet requirement. The RomaNN tag information is ignored for the l-jets. The muon p_T^{rel} templates for charm and light-flavor are shown in Fig. 14.

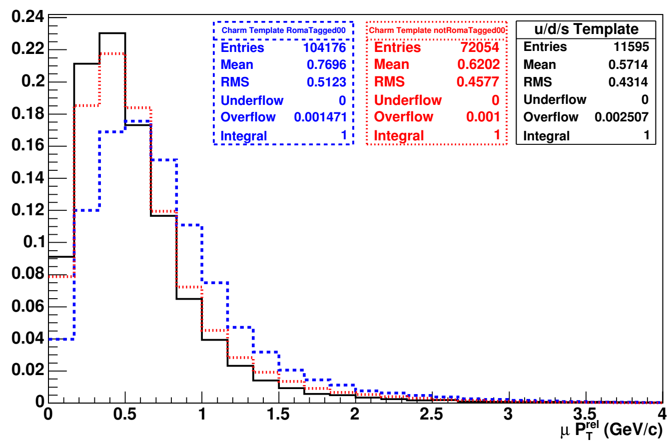


Figure 14: Comparison of the non-b muon p_T^{rel} templates, normalized to unit area.

4.3 Fitting The Data p_T^{rel} Distribution

To get the b content in data we multiply the number of jets by the b fraction obtained from fitting the data p_T^{rel} distribution with the b, c, and l-templates. The b fraction in the sample of jets is f_b such that it maximizes in a two component likelihood fit (Eqn. 6).

$$\mathcal{L} = \prod_{i=0}^{Nbins} \mathbf{P}(n_i, \mu_i) \tag{6}$$

where $\mathbf{P}(a,b)$ is the Poisson probability of observing a events when expecting b; n_i is the number of data jets in bin i; μ_i is the expected number of jets in bin i according to (Eqn. 7)

$$\mu_i = N_{data}\left(f_b T_b^i + f_c T_c^i + (1 - f_b - f_c) T_l^i\right) \tag{7}$$

where N_{data} is the total number of data jets in all bins; f_b is the b-fraction and f_c is the c-fraction; T_x^i is the size of the i^{th} bin of template x normalized to unit area.

We perform separate fits for the samples of tagged jets and not-tagged jets in data, because the b-templates in these two categories have different shapes. In addition, the muon c-templates are split into two categories because they too have different shapes. The electron c-templates are not split into two categories because they have identical shapes. The l-templates are not split into two categories because they have very low statistics.

5 Results for RomaNN3out>0.00

The data efficiency of the RomaNN tagging b-jets has been measured using the lepton p_T^{rel} technique, using both the muon and the electron as the lepton. The MC and data efficiency for tagging b-jets are calculated according to equation (4) and the results then produce a scale factor according to equation (5).

5.1 Generic-Jets as Represented by Electron-Jets

There is an issue that the electron p_T^{rel} spectrum depends on the E_T distribution of the sample (see Figs. 65-97). Fig. 15 shows a comparison between MC b/c/conversion-templates and our data driven b-template (away-jet UltraTightSecVtx tagged), noting that the templates have different E_T distribution thus affecting the p_T^{rel} spectrum. Therefore the fits in the individual E_T bins are more accurate because they are not affected by the difference in E_T distribution between data and MC templates.

Fig. 16 shows a trend of the scale factor as a function of jet E_T. We shift the x-axis Fig. 17 in order to minimize the parameterization error. The linear fit (ignoring the last bin due to insufficient statistics) of the scale factor from 20 GeV to 45 GeV shows a slight slope of 0.00068 ± 0.0024, which is consistant with zero. Therefore we fit a flat line in Fig. 18, to get a scale factor of 0.793 ± 0.018(stat).

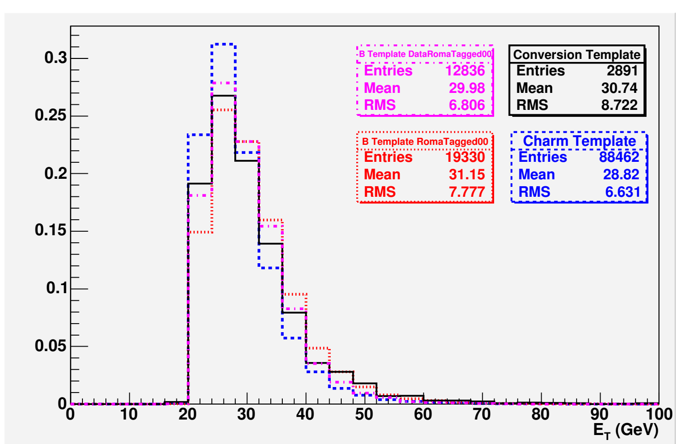


Figure 15: Comparison of the E_T distribution of the jets used to construct our electron p_T^{rel} templates.

Another issue is that the scale factor varies a lot with the number of z vertex in the event, Fig. 19 shows the trend.

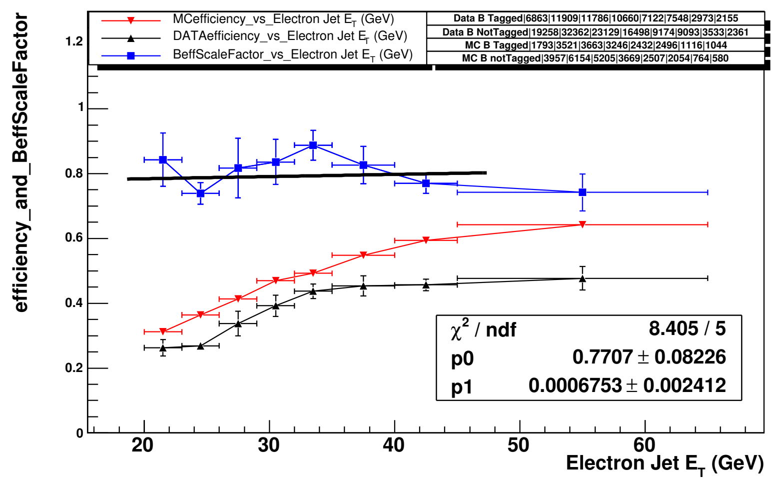


Figure 16: Scale Factor vs. Jet E_T, fitted with line.

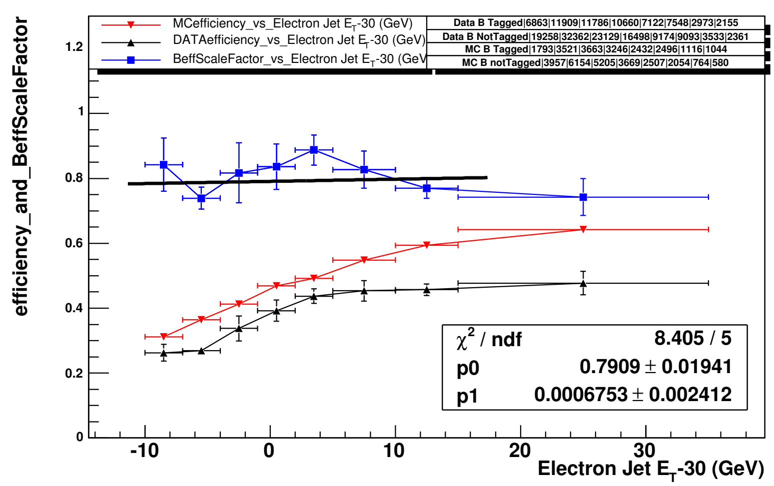


Figure 17: Scale Factor vs. Jet E_T, x-axis shifted to minimize the parameterization error.

The scale factor clearly decreases from about 0.95 for jets with only one z vertex to about 0.75 for jets with only four or more z vertices. We provide a correction factor of $\frac{0.8771-0.02285\times(nZvertex-1.8)}{0.8771}$ (see Fig. 20). The correction factor is centered on 1.8 because it minimizes the parameterization error. It is also the mean number of z vertices for our di-jet data sample. Fig. 21 compares the distribution of the number of z vertex in events from di-jet data, di-jet MC, and TTbar MC.

To get an estimate of the uncertainty of using the parameterization to extrapolate to another sample, we take the combination of the highest/lowest constant value and slopes in Fig. 20, weighted by the TTbar z vertices distribution, and take the standard deviation. The standard deviation is 0.030. Therefore we take 0.030 as the uncertainty of parameterization. In summary, we have a scale factor for generic-jets of
$0.793 \times \frac{0.8771-0.02285\times(nZvertex-1.8)}{0.8771} \pm 0.018(stat) \pm 0.030(nZvertex)$.

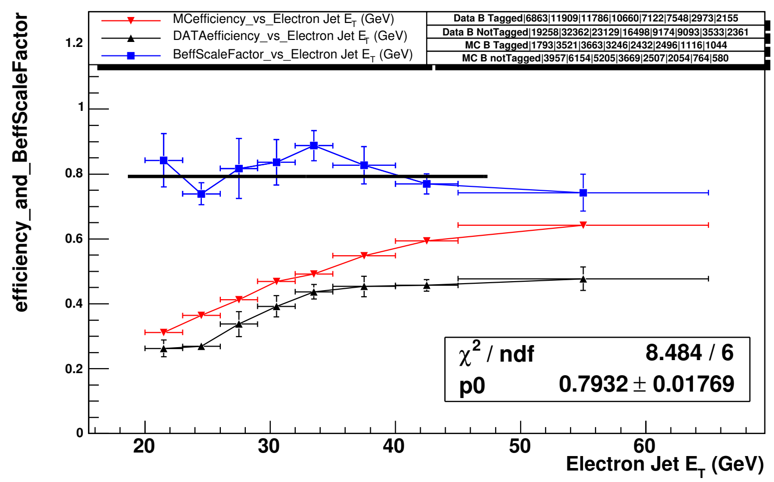


Figure 18: Scale Factor vs. Jet E_T, fitted with flat line.

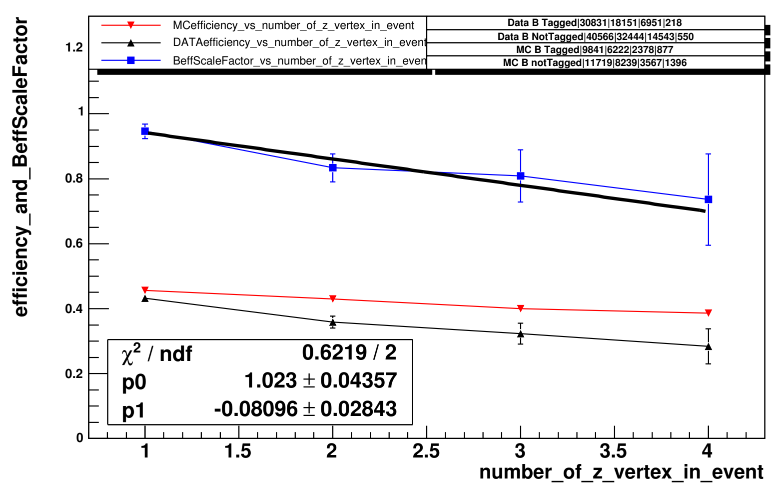


Figure 19: Scale Factor vs. Number Of z Vertex In Event.

5.2 Muon-Jets

Fig. 22 shows a trend of the scale factor as a function of jet E_T, and that the scale factor varies a lot with the transverse energy of the jet.

The scale factor clearly decreases from about 1.05 for jets with $E_T = 20$ GeV, to less than 0.9 for jets with $E_T > 40$ GeV. We provide a parameterization of $(0.8821 - 0.006541 \times (JetE_T(inGeV) - 45))$ (see Fig. 23). The parameterization is centered on 45 GeV because it minimizes the parameterization error. For comparison, the trend is fitted with a flat line in Fig. 24.

Fig. 25 shows the jet E_T spectrum from the di-jet data, di-jet MC, and TTbar MC. It is clear that the b-jets from TTbar have higher average jet E_T than those in the di-jet calibration sample. To get an estimate of the uncertainty of using the parameterization to extrapolate to another sample, we take the combination of the highest/lowest constant value and slopes Fig. 23, weighted by the worst case TTbar E_T spectrum, and take the standard deviation.This takes into account the effect of the parameterization uncertainty, which is a

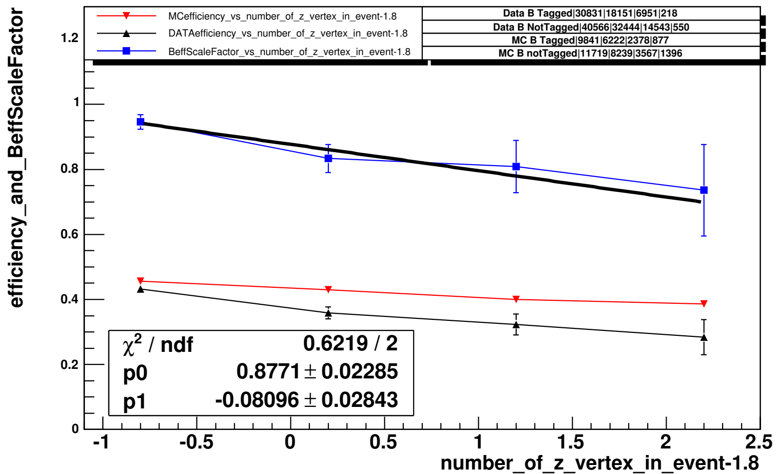


Figure 20: Scale Factor vs. Number Of z Vertex In Event, x-axis shifted to minimize the parameterization error.

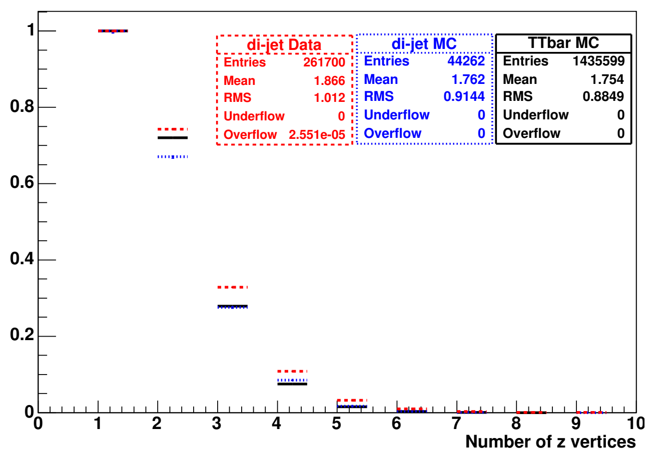


Figure 21: Comparison of the distribution of the number of z vertex in events from di-jet data, di-jet MC, and TTbar MC.

statistical error, as well as the systematic error in extrapolating to another sample. The standard deviation is 0.061. Therefore we take 0.061 as the uncertainty of parameterization. In summary, we have a scale factor for muon-jets of
$(0.8821 - 0.006541 \times (JetE_T(inGeV) - 45)) \pm 0.061(stat + JetE_T\ param)$

6 Systematic Errors and Cross Checks

6.1 Electron p_T^{rel}

The systematic errors in the scale factor measurement of the binary mode RomaNN for electron-jets are listed in table 12.

For those systematics that are estimated using pseudoexperiments, we construct pseudo-data in the alternate model and always fit using our default model. We use input fractions as obtained in the central value fits (Fig. 52). For tagged, we use input fractions of 84.6%

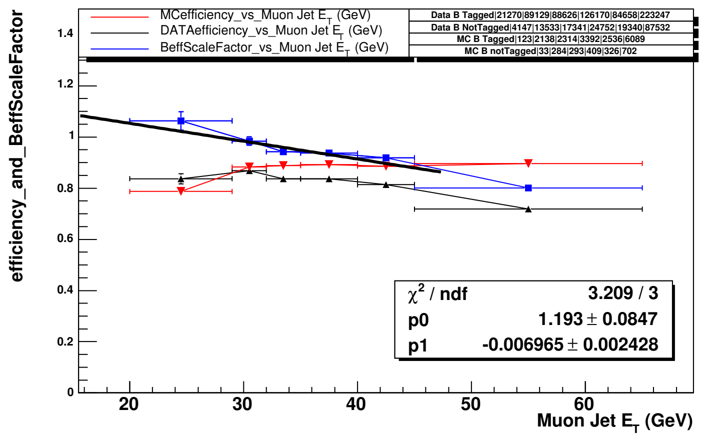


Figure 22: Scale Factor vs. Jet E_T.

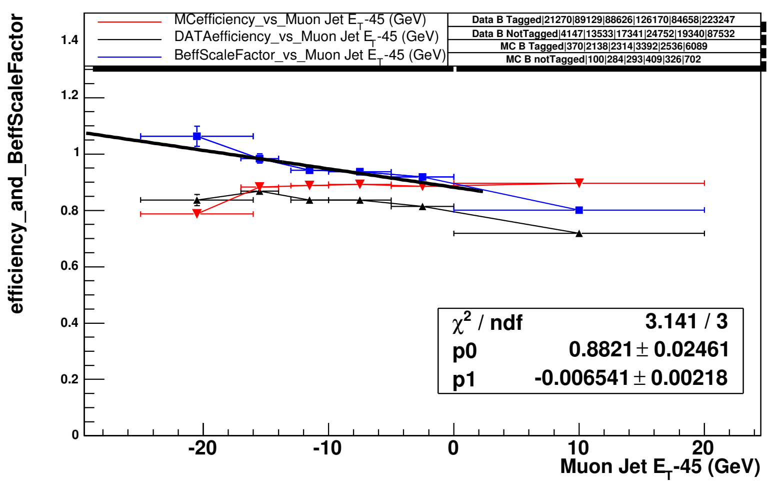


Figure 23: Scale Factor vs. Jet E_T, x-axis shifted to minimize the parameterization error.

b, 4.3% c, and 11.1% conversion. For not-tagged, we use input fractions of 49.4% b, 19.2% c, and 31.4% conversion. We use 10000 events for both tagged and not-tagged, and perform 5000 pseudoexperiments each. The important thing is to determine the shift in b-fraction. 10000 events is enough to fill a distribution from templates which had 18 bins. 5000 pseudo-experiments is enough such that the mean of the fitted result has only a 1% error. We use the number of tagged and not-tagged events in data and only change the b-fraction according to the systematic shift found.

We will compare the distribution of various quantities from the di-jet data, di-jet MC, and TTbar MC. For di-jet data, we are showing only jets with electrons. For di-jet MC, we are showing only b-jets with electrons. For TTbar we are showing every b-jets.

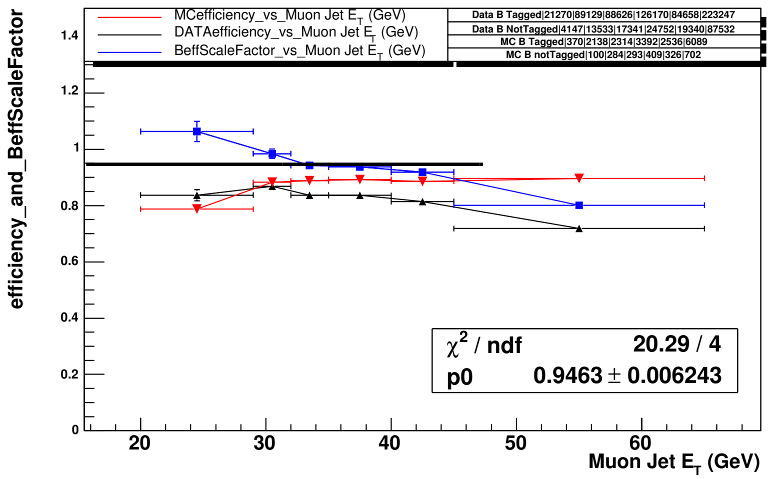


Figure 24: Scale Factor vs. Jet E_T, fitted with flat line.

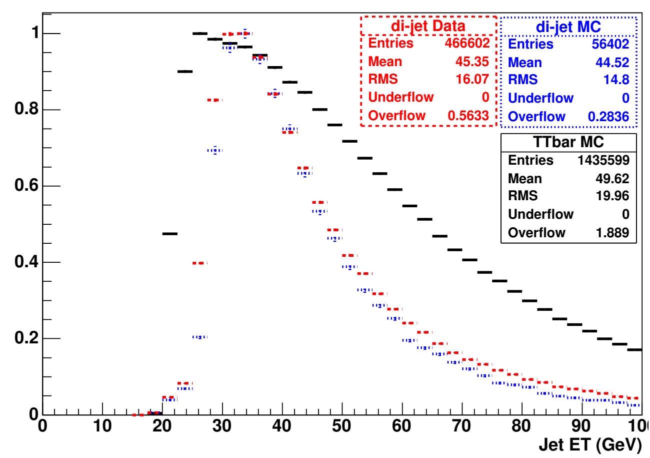


Figure 25: Comparison of the E_T distribution of muon-jets from di-jet data, di-jet MC, and TTbar MC.

	Source	*Systematic Error*
-	Jet Transverse Energy	0.051
-	Track Multiplicity	0.006
-	Conversion Model	0.019
-	MC b-jet Model	0.058
	Total	0.080

Table 12: Systematic errors in RomaNN scale factor measurement.

6.1.1 Jet Transverse Energy

The b jets from other physics processes might have an E_T distribution different from the sample used to measure the scale factor. Fig. 26 shows the jet E_T spectrum from the di-jet data, di-jet MC, and TTbar MC. It is clear that the b-jets from TTbar have higher average jet E_T than those in the di-jet calibration sample. A worst case scenerio using the TTbar sample will be used to weight the bins to estimate a systematic error.

As a systematic error, we look at the worst case scenerio of extrapolating to the TTbar sample. We weight the E_T spectrum of the TTbar MC against the linear fit with plus and minus standard deviation of slope error in Fig. 17, and look at the range/2. The scale factor weighted with $+1\sigma$ error is 0.754, with $+0\sigma$ error is 0.805, with -1σ error is 0.856, which has a $\frac{range}{2}$ of 0.051.

$$SF_{Weighted} = \sum_{all\ bins} (p0 - (p1 \times BinCentroid)) \frac{BinWeight}{\sum_{all\ bins} (BinWeight)} \tag{8}$$

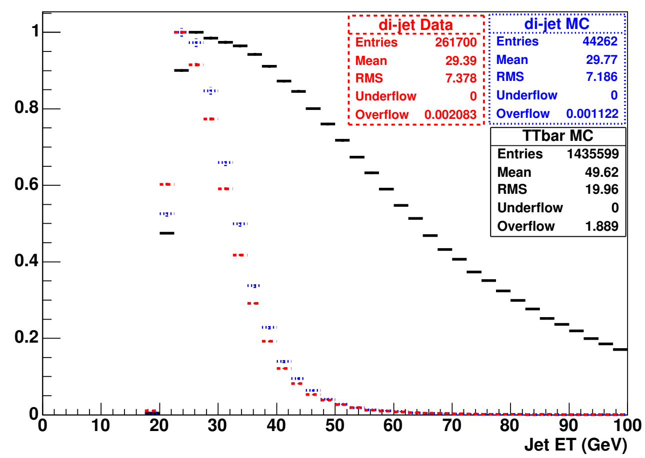


Figure 26: Comparison of the E_T distribution of electron-jets from di-jet data, di-jet MC, and TTbar MC.

6.1.2 Track Multiplicity

The scale factor is measured on a sample of b-jets where the B hadrons decayed semileptonically. If the scale factor depends on the track multiplicity, then it may be different for the hadronic decays of the B hadrons. Fig. 27 compares the distribution of the number of Roma Good Tracks inside the electron-jets from di-jet data, di-jet MC, and TTbar MC. Fig. 28 shows a trend of the scale factor as a function of the number of RomaNN good tracks embedded in the jet. A RomaNN good track is defined in Table 13. A worst case scenerio using the TTbar sample will be used to weight the bins to estimate a systematic error.

As a systematic error, we look at the worst case scenerio of extrapolating to the TTbar sample. We weight the $nTrack$ spectrum (Fig. 28) bin by bin with the di-jet MC and the TTbar MC, and compare their weighted scale factor against each other. The scale factor weighted with di-jet MC is 0.875, TTbar is 0.881, which is a shift of 0.006.

The fits for each individual point can be found in the appendix.

Requirements
collection == DefTrack
$\lvert\Delta R(track, jet)\rvert < 0.4$
$P_T > 1$ GeV/c
$z_0 < 1.0$ cm
$d_0 < 0.15$ cm
Reconstructed with both Silicon and COT hits

Table 13: RomaNN good track requirements.

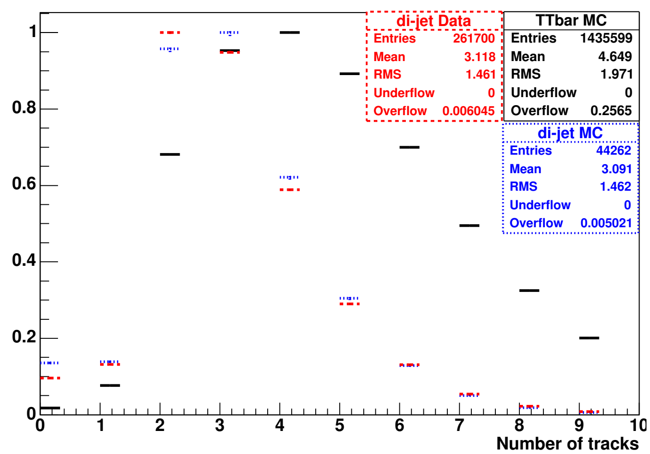


Figure 27: Comparison of the distribution of the number of Roma Good Tracks inside the electron-jets from di-jet data, di-jet MC, and TTbar MC.

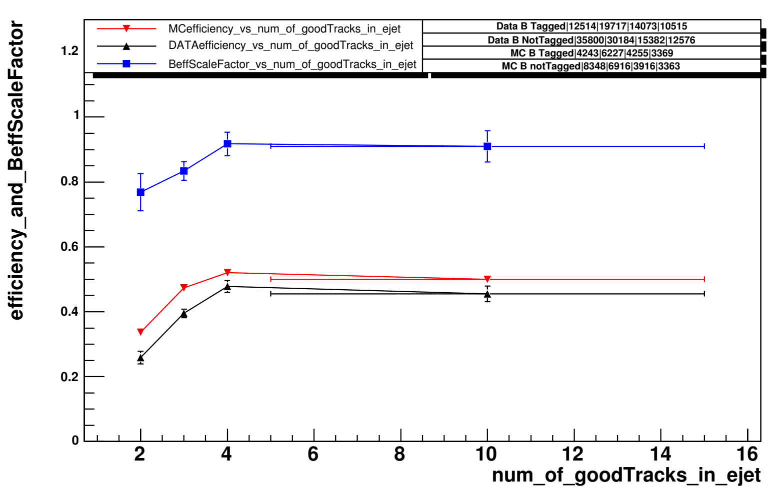


Figure 28: Scale Factor vs. Number of RomaNN good tracks.

6.1.3 Conversion Template Shape

We estimate how sensitive the scale factor is to the choice of conversion template. The default light-flavor templates are constructed from data by requiring the away-jet to be negatively tagged by Loose SecVtx, and the electron inside the electron-jet to have 0 silicon hits registered where there should have been hits. Figure 29 shows the distribution of number of the silicon hits for the electron in the data sample.

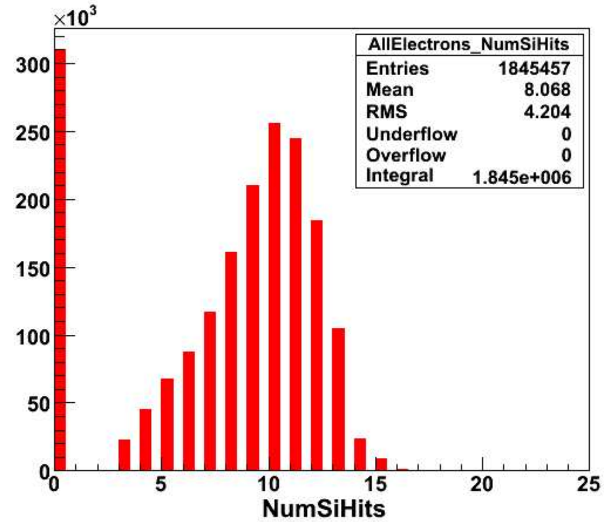


Figure 29: The distribution of number of Si-Hits in the electron-jet in the di-jet data sample.

We build 3 other light-flavor templates with 3/4/5 silicon hits registered instead of zero silicon hits registered. The four l-templates along with the default b and c-templates were used to generate 4 sets of pseudodata. These 4 sets of pseudodata are then fitted with the default templates. The shift in scale factor would then be a good estimate of the error caused by relying on a particular method of l-template construction. The standard deviation of the scale factor is 0.019, so 0.019 is taken as the systematic uncertainty in this category.

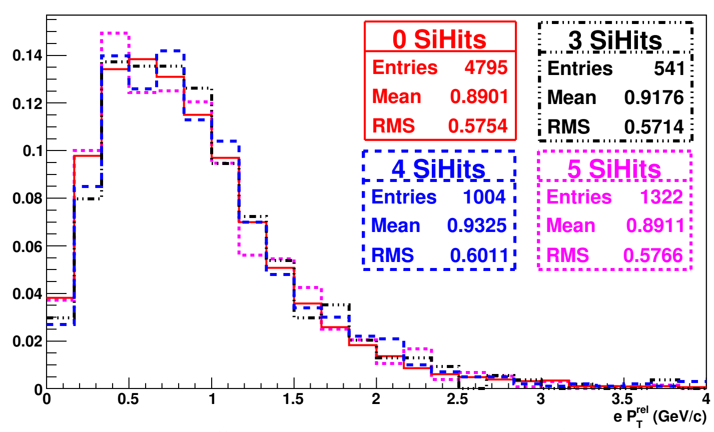


Figure 30: 4 different l-templates, built from data.

6.1.4 MC b-jet Model

We need to estimate how sensitive the scale factor is to our model of electron-b-jets. We have two method of studying this: one based on comparison between data and MC, another

Conversion Template	*Tagged* $< f_b^{fitted} >$	*Tagged Fit mean* χ^2/DOF	*Not Tagged* $< f_b^{fitted} >$	*Not Tagged Fit mean* χ^2/DOF	*scalefactor*
Zero Si-hits	0.846	1.127	0.493	1.134	0.890
Three Si-hits	0.866	1.329	0.542	4.119	0.851
Four Si-hits	0.860	1.219	0.524	2.124	0.865
Five Si-hits	0.866	1.273	0.544	2.605	0.849

Table 14: Dependence of RomaNN scale factor from the Conversion Model used.

one based on MC-tuning.

The first study we estimate the scale factor sensitivity to the difference in our MC model of b-jets and an alternative data model of b-jets.

For the data based study, we obtain a data sample very enriched in b-jets by requiring the away-jet to be tagged by the UltraTight SecVtx tagger, and that the electron-jet be tagged with a RomaNNcut=0.8. From Fig. 3, we can assume that the RomaNN tagged electron-jets in this sample are all from b-jets, with no contamination from charm or light. We use the p_T^{rel} of these tagged electron-jets as an alternate model of tagged b-jets.

We construct pseudoexperiments using the alternate model for tagged b-jets (with a RomaNNcut=0.8) and the default models for charm, conversion, and not-tagged b-jets. We fit using the default model for tagged b-jets (with a RomaNNcut=0.8) and the default models for charm, conversion, and not-tagged b-jets. We find a b-fraction of 0.766 instead of the input 0.846. When we propagate this to the scale factor we get 0.836, which is a shift of 0.050. We take this shift to be the systematic uncertainty estimation using this data based method.

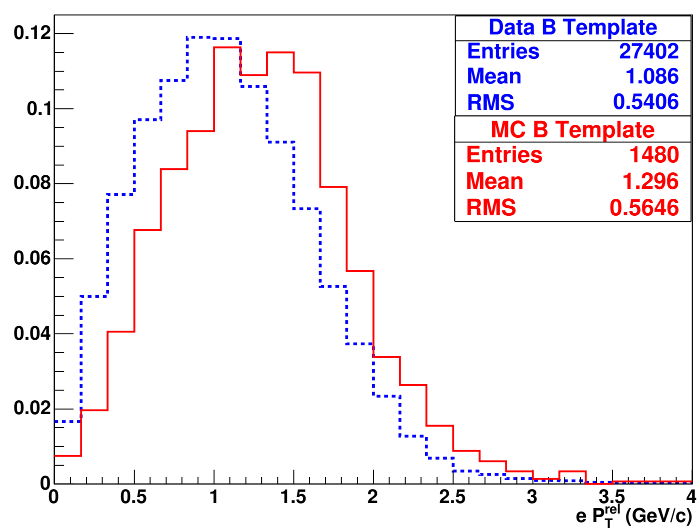


Figure 31: Comparison of the tagged templates, red = MC based(default) b-template, blue = data based b-template.

The second study we estimate the scale factor sensitivity to the MC-tuning.

The model of b-jets depends on an accurate simulation of heavy flavor hadron production and decay. The fragmentation model is the mechanism responsible for assigning in MC the b-quark's momentum amongst the B hadron and the other hadron in the b-jet. The fragmentation model used in the default templates is the Bowler-Lund model, which uses parameters as fitted from OPAL [11].

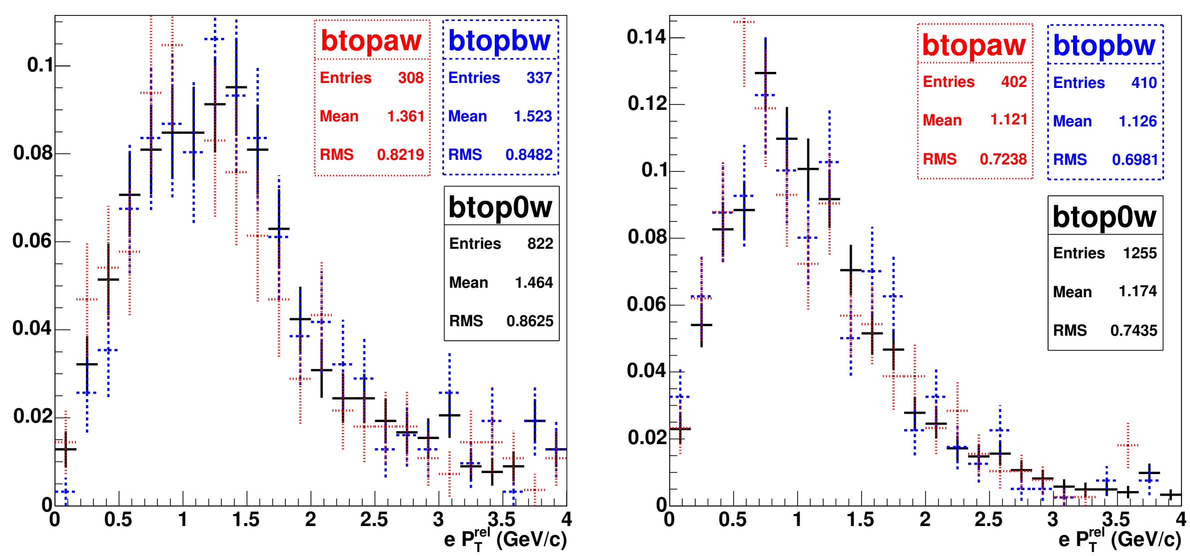
In order to investigate the dependence of the scale factor to the fragmentation model used in the MC simulations, we would ideally use MCs with Bowler-Lund within a band of its parameters' uncertainty, and accordingly and perform pseudoexperiments. But we only have one MC generated by the Bowler-Lund model.

So the strategy used is to estimate the scale factor shift within a band of parameters' uncertainty of an alternative fragmentation model, the Peterson model, of which we fortunately have 2 existing $W + b\bar{b}$ MC. These two MCs used PFP=(0.0025,0.0041), the lower bound of ALEPH (0.0031 ± 0.0006) and the central value of OPAL (0.00412 ± 0.00037) respectively [11]. Then we would have the scale factor shift as a function of PFP shift, with which we can get the scale factor uncertainty as a result of PFP uncertainty. We would use the PFP uncertainty from OPAL since the Bowler-Lund model used in our scale factor value calculation uses parameters as fitted from OPAL.

We also have a $W + b\bar{b}$ MC based on the Bowler-Lund model. The 3 $W + b\bar{b}$ MCs were used to build 3 different b-templates (Fig. 32), these along with the default c and l-templates were used to generate 3 sets of pseudodata. These 3 sets of pseudodata are then fitted with the b-template from the Bowler-Lund modeled $W + b\bar{b}$, along with the default c and l-templates. We must fit using the b-template from the Bowler-Lund modeled $W + b\bar{b}$ instead of the di-jet b-templates since we generated the Peterson pseudodata with b-templates from $W + b\bar{b}$ MC.

A shift in PFP of 0.0041-0.0025=0.0016 changes the scale factor by 1.049-0.799=0.250, so the scale factor uncertainty as a result of PFP uncertainty would be $0.00037 \times \frac{0.250}{0.0016} = 0.058$. We take this shift of 0.058 to be the systematic uncertainty estimation using this MC-tuning based method.

To be conservative, we take the larger of the two errors (table 16) to be the estimate of how sensitive the scale factor is to our model of electron-b-jets, and assign a systematic error of 0.058.



(a) Tagged templates. (b) Not Tagged templates.

Figure 32: 3 different b-templates, built from $W + b\bar{b}$ MCs. Note that jets from $W + b\bar{b}$ MC have higher E_T than from di-jet MC, so that the electron p_T^{rel} is shifted to higher values (in the tagged templates: 1.46 for Bowler-Lund $W + b\bar{b}$ compared to 1.13 for Bowler-Lund di-jet (Fig. 11)).

Fragmentation Model	*Tagged* $< f_b^{fitted} >$	*Tagged Fit mean* χ^2/DOF	*Not Tagged* $< f_b^{fitted} >$	*Not Tagged Fit mean* χ^2/DOF	*scale factor*
Bowler-Lund	0.846	1.005	0.493	1.010	0.889
Peterson PFP=0.0025	0.765	59.0	0.528	47.7	0.799
Peterson PFP=0.0041	0.978	17.6	0.428	20.0	1.049

Table 15: Dependence of RomaNN scale factor from the Fragmentation Model used.

Method of Estimation	*Estimated Value*	*Assigned Value*
Data b-model	0.050	no
Tuned MC b-model	0.058	yes

Table 16: Summary of the estimation in the MC b-jet Model systematic error.

6.1.5 Jet η

Fig. 33 compares the η distribution of the electron-jets from di-jet data, di-jet MC, and TTbar MC. As a cross-check we show that the scale factor varies minimally within the η range of our measurement in Fig. 34. We do not quote a systematic error for this.

The fits for each individual point can be found in the appendix.

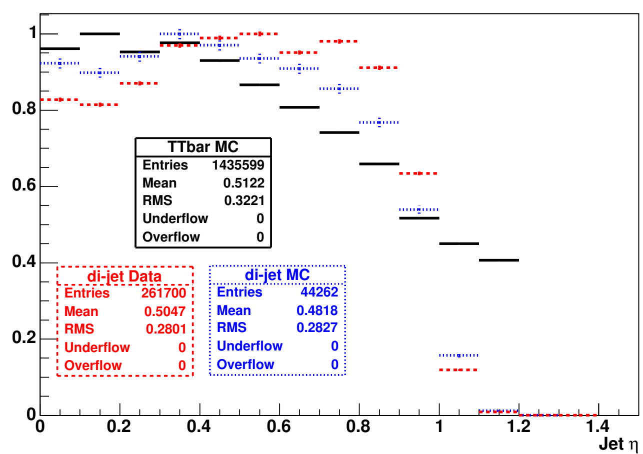


Figure 33: Comparison of the η distribution of electron-jets from di-jet data, di-jet MC, and TTbar MC.

6.1.6 Jet ϕ

Fig. 35 compares the ϕ distribution of the electron-jets from di-jet data, di-jet MC, and TTbar MC. As a cross-check we show that the scale factor varies minimally within the ϕ range of our measurement in Fig. 36. We do not quote a systematic error for this.

The fits for each individual point can be found in the appendix.

6.1.7 Data Period

Fig. 37 shows the distribution of the number of events in the runs used. Fig. 38 shows that the scale factor varies a lot with the data period. This is correlated to the fact that the

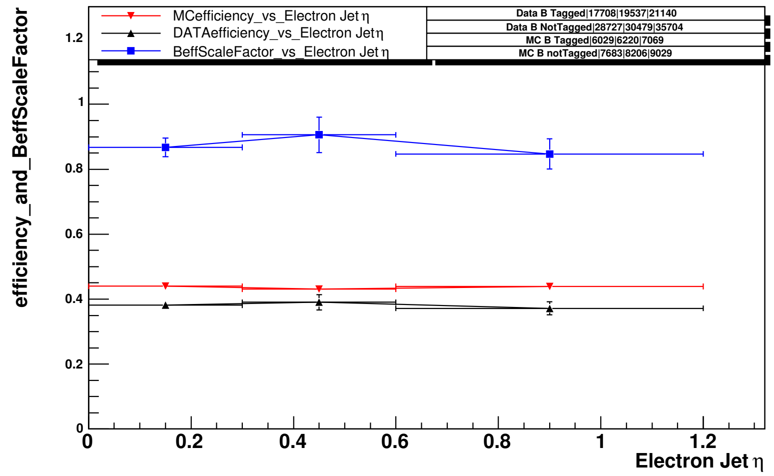


Figure 34: Scale Factor vs. Jet Eta.

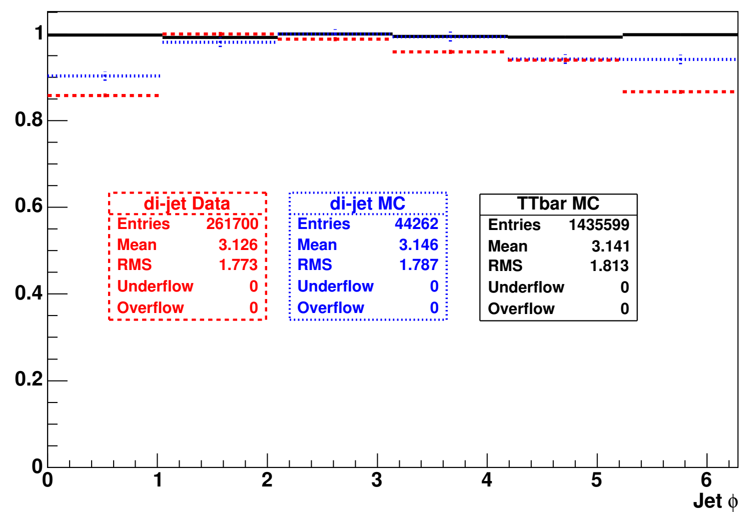


Figure 35: Comparison of the ϕ distribution of electron-jets from di-jet data, di-jet MC, and TTbar MC.

instaneous luminosity is different for each data period, so we do not quote another systematic error.

The fits for each individual point can be found in the appendix.

6.1.8 Even and Odd Event Number

As a cross-check we show that the scale factor varies minimally between odd and even events in Fig. 39. We do not quote a systematic error for this.

The fits for each individual point can be found in the appendix.

6.2 Muon p_T^{rel}

Muon p_T^{rel} is a well tested method, so it is just summerized in table 17 and not described in too much detail, leaving more room for the Electron p_T^{rel} method instead. Trends are

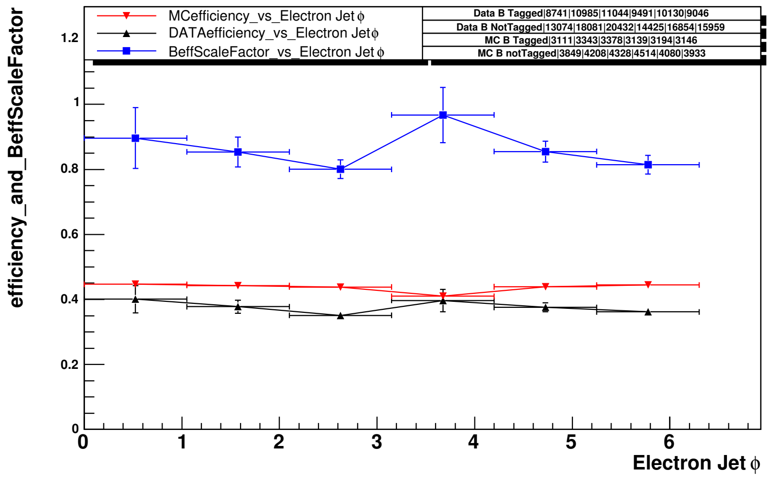


Figure 36: Scale Factor vs. Jet Phi.

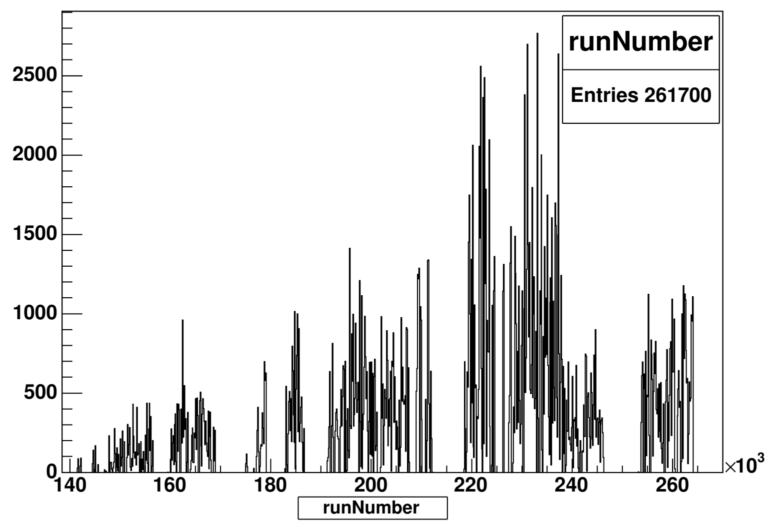


Figure 37: Distribution of the data over run number from di-jet data.

provided here to build confidence in this measurement.

6.2.1 Track Multiplicity

Fig. 40 compares the distribution of the number of Roma Good Tracks inside the muon-jets from di-jet data, di-jet MC, and TTbar MC. Fig. 41 shows a trend of the scale factor as a function of the number of RomaNN good tracks embedded in the jet.

As a systematic error, we look at the worst case scenerio of extrapolating to the TTbar sample. We weight the *nTrack* spectrum (Fig. 28) bin by bin with the di-jet MC and the TTbar MC, and compare their weighted scale factor against each other. The scale factor weighted with di-jet MC is 0.915, TTbar is 0.893, which is a shift of 0.022, to be taken as the systematic error.

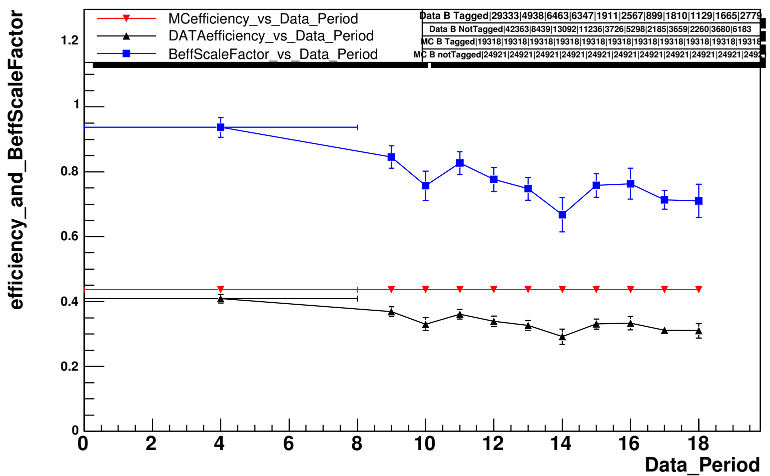


Figure 38: Scale Factor vs. Data Period.

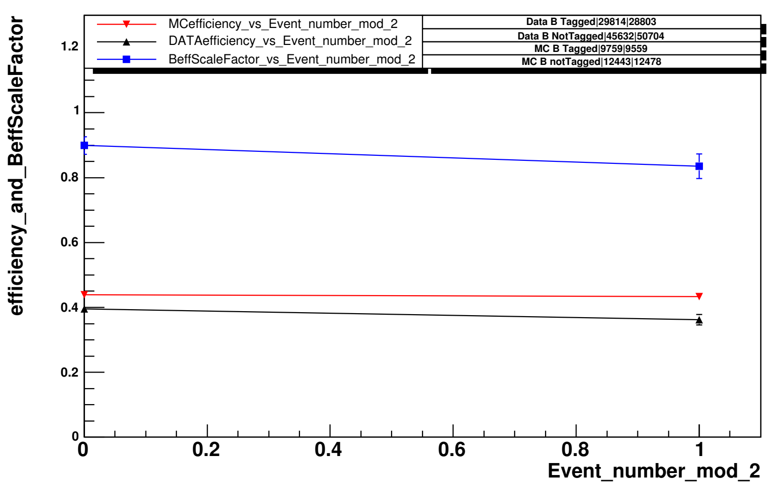


Figure 39: Scale Factor vs. Even and Odd Event Number.

	Source	*Systematic Error*
-	Track Multiplicity	0.022
-	MC b-jet Model	0.018
	Total	0.028

Table 17: Systematic errors in RomaNN scale factor measurement for muon-jets.

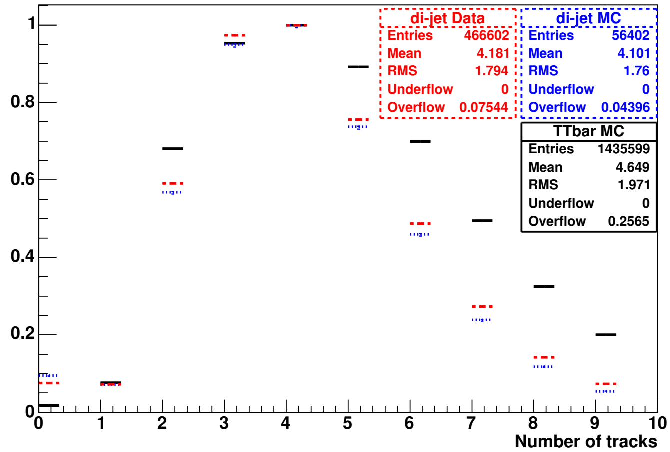


Figure 40: Comparison of the distribution of the number of Roma Good Tracks inside the muon-jets from di-jet data, di-jet MC, and TTbar MC.

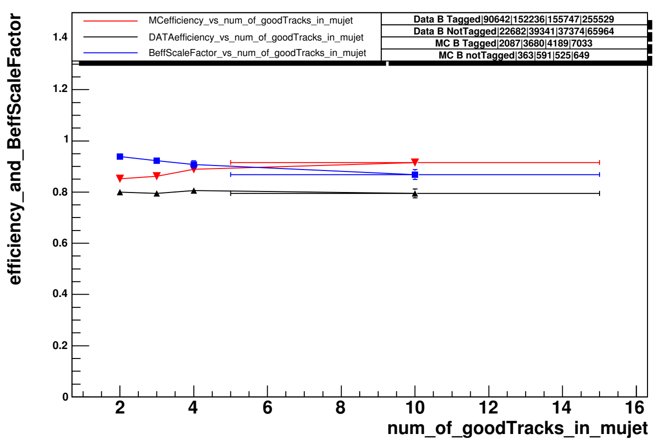


Figure 41: Scale Factor vs. Number of RomaNN good tracks.

6.2.2 MC b-jet Model

We need to estimate how sensitive the scale factor is to our model of muon-b-jets. We have two method of studying this: one based on comparison between data and MC, another one based on MC-tuning.

The first study we estimate the scale factor sensitivity to the difference in our MC model of b-jets and an alternative data model of b-jets.

For the data based study, we obtain a data sample very enriched in b-jets by requiring the away-jet to be tagged by the UltraTight SecVtx tagger, and that the electron-jet be tagged with a RomaNNcut=0.8. From Fig. 3, we can assume that the RomaNN tagged electron-jets in this sample are all from b-jets, with no contamination from charm or light. We use the p_T^{rel} of these tagged electron-jets as an alternate model of tagged b-jets.

We construct pseudoexperiments using the alternate model for tagged b-jets (with a RomaNNcut=0.8) and the default models for charm, conversion, and not-tagged b-jets. We fit using the default model for tagged b-jets (with a RomaNNcut=0.8) and the default models for charm, conversion, and not-tagged b-jets. We find a b-fraction of 0.630 instead of the

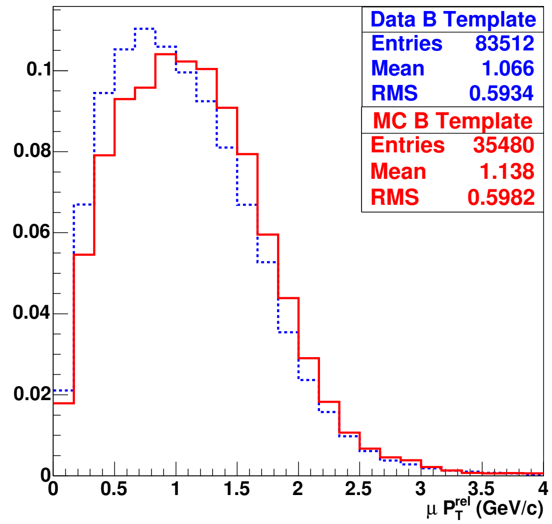


Figure 42: Comparison of the tagged templates, red = MC based(default) b-template, blue = data based b-template.

input 0.698. When we propagate this to the scale factor we get 0.875, which is a shift of 0.015. We take this shift to be the systematic uncertainty estimation using this data based method.

The second study we estimate the scale factor sensitivity to the MC-tuning.

Like for the electron-jets, the strategy used is to estimate the scale factor shift within a band of parameters' uncertainty of an alternative fragmentation model, the Peterson model, of which we fortunately have 2 existing $W+b\bar{b}$ MC. These two MCs used PFP=(0.0025,0.0041), the lower bound of ALEPH (0.0031 ± 0.0006) and the central value of OPAL (0.00412 ± 0.00037) respectively [11]. Then we would have the scale factor shift as a function of PFP shift, with which we can get the scale factor uncertainty as a result of PFP uncertainty. We would use the PFP uncertainty from OPAL since the Bowler-Lund model used in our scale factor value calculation uses parameters as fitted from OPAL.

We also have a $W + b\bar{b}$ MC based on the Bowler-Lund model. The 3 $W + b\bar{b}$ MCs were used to build 3 different b-templates (Fig. 32), these along with the default c and l-templates were used to generate 3 sets of pseudodata. These 3 sets of pseudodata are then fitted with the b-template from the Bowler-Lund modeled $W + b\bar{b}$, along with the default c and l-templates. We must fit using the b-template from the Bowler-Lund modeled $W+b\bar{b}$ instead of the di-jet b-templates since we generated the Peterson pseudodata with b-templates from $W + b\bar{b}$ MC.

A shift in PFP of 0.0041-0.0025=0.0016 changes the scale factor by 0.998-0.919=0.079, so the scale factor uncertainty as a result of PFP uncertainty would be $0.00037 \times \frac{0.079}{0.0016} = 0.018$.

To be conservative, we take the larger of the two errors (table 19) to be the estimate of how sensitive the scale factor is to our model of muon-b-jets, and assign a systematic error of 0.018.

6.2.3 Number Of z Vertex In Event

Fig. 43 compares the distribution of the number of z vertex in events from di-jet data, di-jet MC, and TTbar MC. Fig. 44 shows a trend of the scale factor as a function of the number of z vertex in the event. We do not quote a systematic error for this.

Fragmentation Model	*Tagged* $< f_b^{fitted} >$	*Tagged Fit mean* χ^2/DOF	*Not Tagged* $< f_b^{fitted} >$	*Not Tagged Fit mean* χ^2/DOF	*scale factor*
Bowler-Lund	0.697	1.02	0.318	1.03	0.894
Peterson PFP=0.0025	0.503	1.68	0.198	4.37	0.919
Peterson PFP=0.0041	0.648	1.72	0.145	2.58	0.998

Table 18: Dependence of RomaNN scale factor from the Fragmentation Model used.

Method of Estimation	*Estimated Value*	*Assigned Value*
Data b-model	0.015	no
Tuned MC b-model	0.018	yes

Table 19: Summary of the estimation in the MC b-jet Model systematic error.

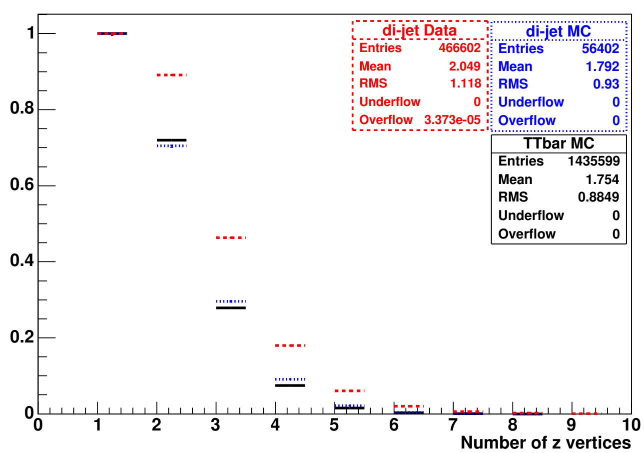


Figure 43: Comparison of the distribution of the number of z vertex in events from di-jet data, di-jet MC, and TTbar MC.

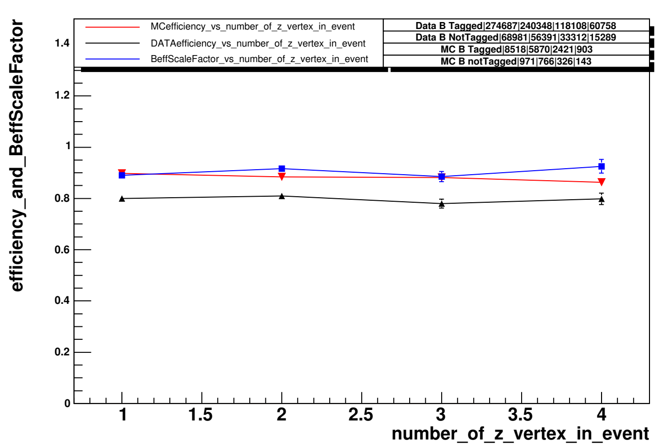


Figure 44: Scale Factor vs. Number Of z Vertex In Event.

6.2.4 Jet η

Fig. 33 compares the η distribution of the muon-jets from di-jet data, di-jet MC, and TTbar MC. As a cross-check we show that the scale factor varies minimally within the η range of our measurement in Fig. 46. We do not quote a systematic error for this.

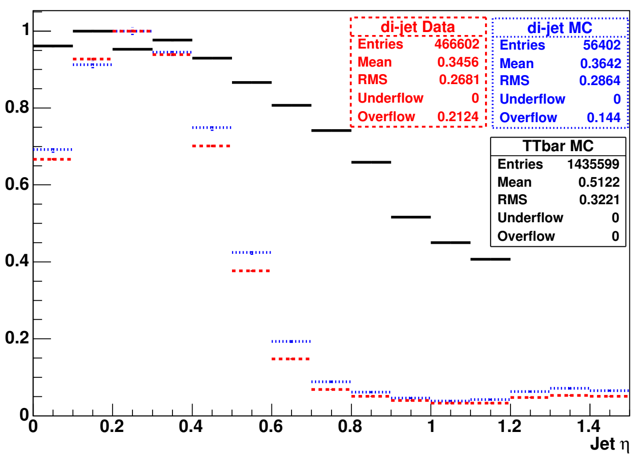


Figure 45: Comparison of the η distribution of electron-jets from di-jet data, di-jet MC, and TTbar MC.

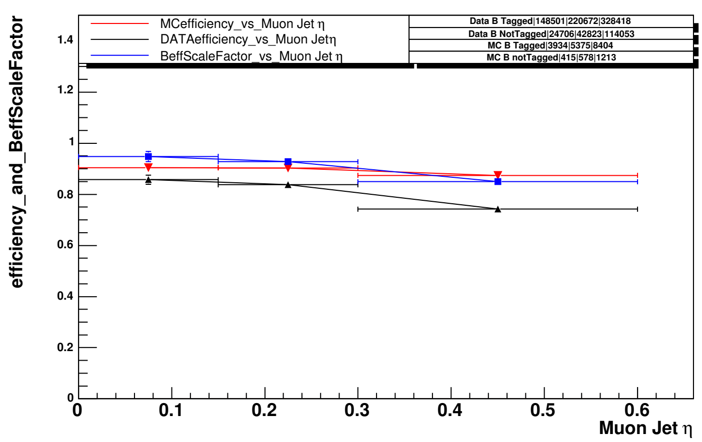


Figure 46: Scale Factor vs. Jet Eta.

6.2.5 Jet ϕ

Fig. 47 compares the ϕ distribution of the electron-jets from di-jet data, di-jet MC, and TTbar MC. As a cross-check we show that the scale factor varies minimally within the ϕ range of our measurement in Fig. 48. We do not quote a systematic error for this.

6.2.6 Data Period

Fig. 49 shows the distribution of the number of events in the runs used. Fig. 50 shows a trend of the scale factor as a function of the data period. We do not quote a systematic error for this.

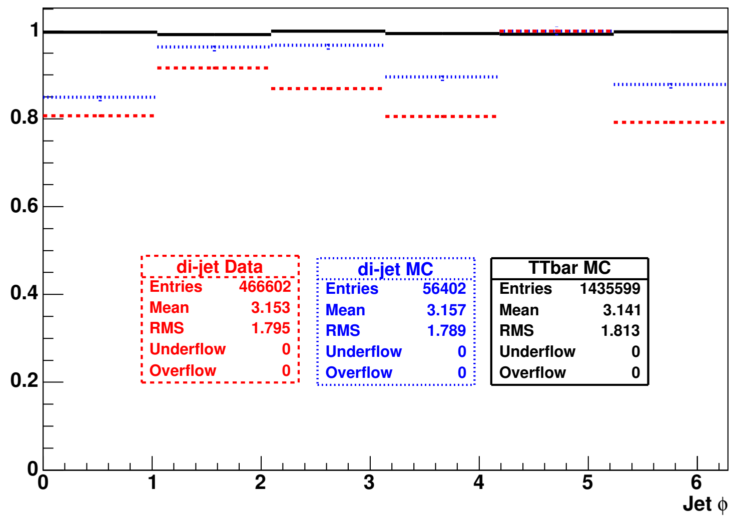


Figure 47: Comparison of the ϕ distribution of electron-jets from di-jet data, di-jet MC, and TTbar MC.

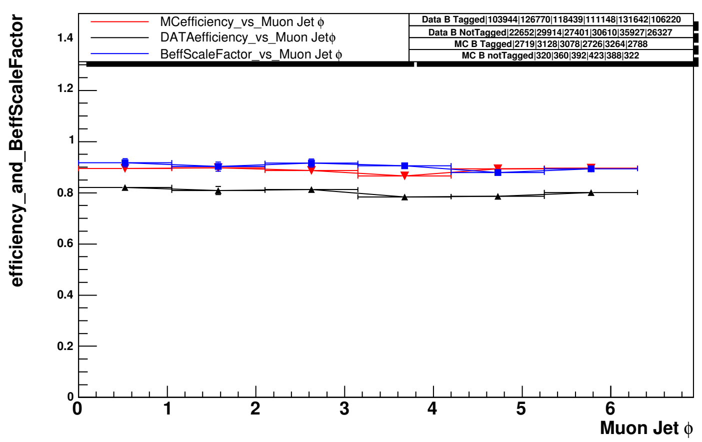


Figure 48: Scale Factor vs. Jet Phi.

6.2.7 Even and Odd Event Number

As a cross-check we show that the scale factor varies minimally between odd and even events in Fig. 51. We do not quote a systematic error for this.

7 Conclusion

The efficiency and scale factor measurement is performed for the binary mode RomaNN using the electron p_T^{rel} technique. The operating point RomaNN3out>0.00, has a scale factor of $(0.8821 - 0.006541 \times (JetE_T(inGeV) - 45)) \pm 0.061(stat + JetE_T\ param) \pm 0.028(syst)$ for all muon jets, and for all other jets a scale factor of
$0.793 \times \frac{0.8771-0.02285\times(nZvertex-1.8)}{0.8771} \pm 0.018(stat) \pm 0.030(nZvertex) \pm 0.080(syst)$.

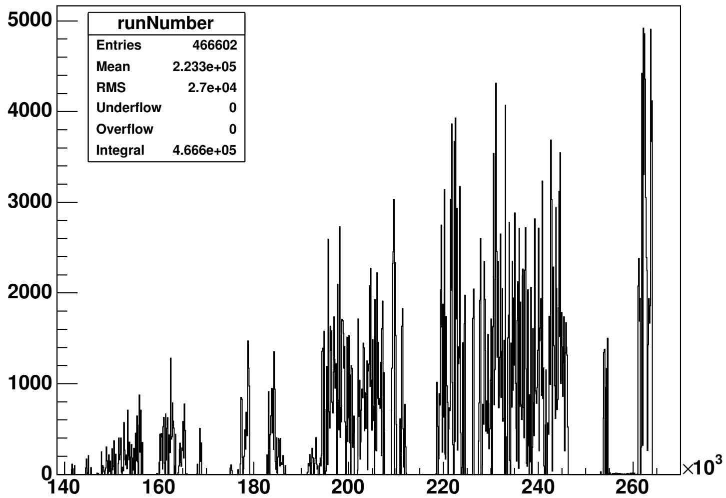


Figure 49: Distribution of the data over run number from di-jet data.

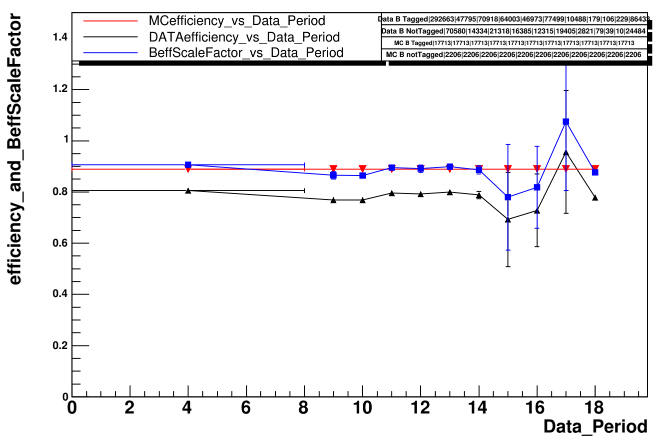


Figure 50: Scale Factor vs. Data Period.

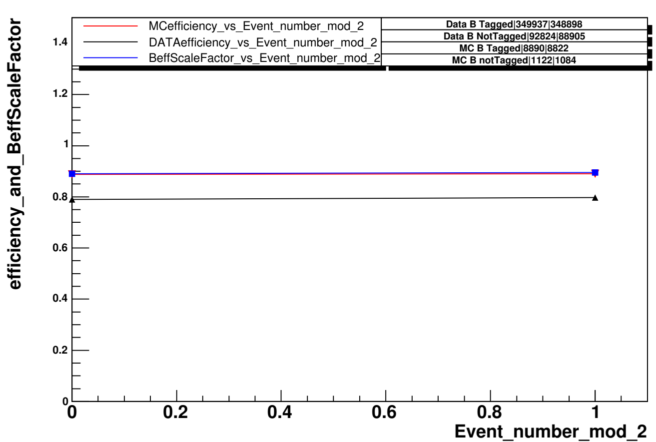


Figure 51: Scale Factor vs. Even and Odd Event Number.

8 Appendix

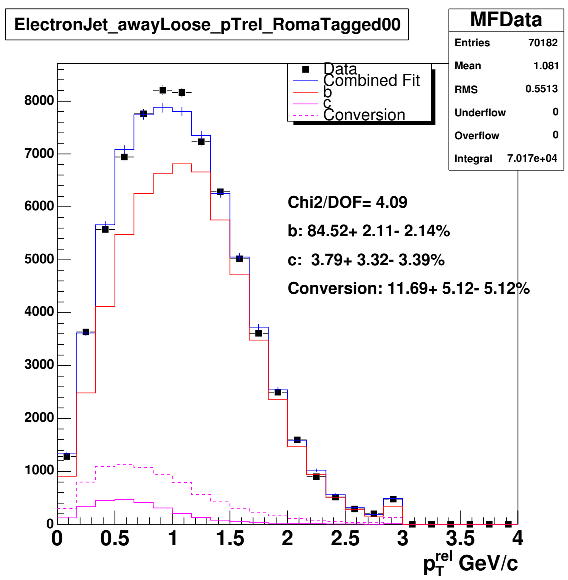
8.1 Naive Fits and Calculations Using Entire Data Sample

8.1.1 Electron p_T^{rel}

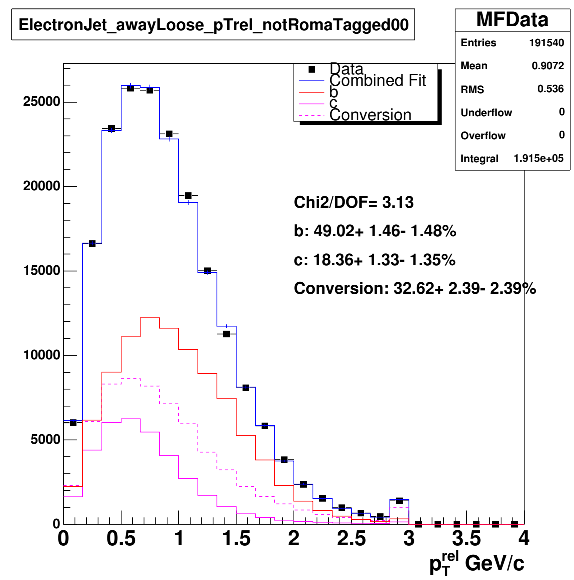
Using the definition of RomaNN3out>0.0 for tagged, the data distribution and best fit results are shown in Fig. 52a for tagged jets and Fig. 52b for not-tagged jets. Table 20 summarizes the best fit results.

With 70,182 tagged jets in data, the best fit finds a b-fraction of $84.5 \pm 2.2\%$, a charm-fraction of $3.8 \pm 3.4\%$, and a light/conversion-fraction of $11.7 \pm 5.1\%$, with 18 bins of 3/18 GeV/c and 2 fit parameters (f_b and f_c). The χ^2/degree-of-freedom is 4.1.

With 191,540 not-tagged jets in data, the best fit finds a b-fraction of $49.0{\pm}1.5\%$, a charm-fraction of $18.4{\pm}1.4\%$, and a light/conversion-fraction of $32.6{\pm}2.4\%$. The not-tagged sample has a lower b-fraction than the tagged sample as expected. The χ^2/degree-of-freedom is 3.1.



(a) Tagged, total # of jets = 70182, b fraction = 84.5%



(b) Not Tagged, total # of jets = 191540, b fraction = 49.0%

Figure 52: Fitting b and non-b components to elec p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged.

The data efficiency for this sample of jets is then

$$\epsilon_{data} = \frac{0.845 \times 70182}{0.846 \times 70182 + 0.490 \times 191540} = \frac{59308}{59308 + 93885} = 0.387 \pm 0.009 \tag{9}$$

In MC, the efficiency is estimated from the number of OBSP b-jets (passing the same kinematic cuts)

$$\epsilon_{MC} = \frac{19330}{19330 + 24932} = 0.437 \pm 0.002 \tag{10}$$

Therefore, the naive scale factor is

$$SF_{elec} = \frac{\epsilon_{data}}{\epsilon_{MC}} = \frac{0.387}{0.437} = 0.886 \pm 0.025 \tag{11}$$

However, this central value cannot be taken as the official scale factor, since the electron p_T^{rel} spectrum depends on the E_T distribution of the sample. This is discussed in section 6.1.1.

Sample	*#tagged*	*%b*	*#b tagged*	*#not-tagged*	*%b*	*#b not-tagged*	ϵ_b
MC	19330	n/a	19330	24932	n/a	24932	0.437
data	70182	84.5	59308	191540	49.0	93885	0.387

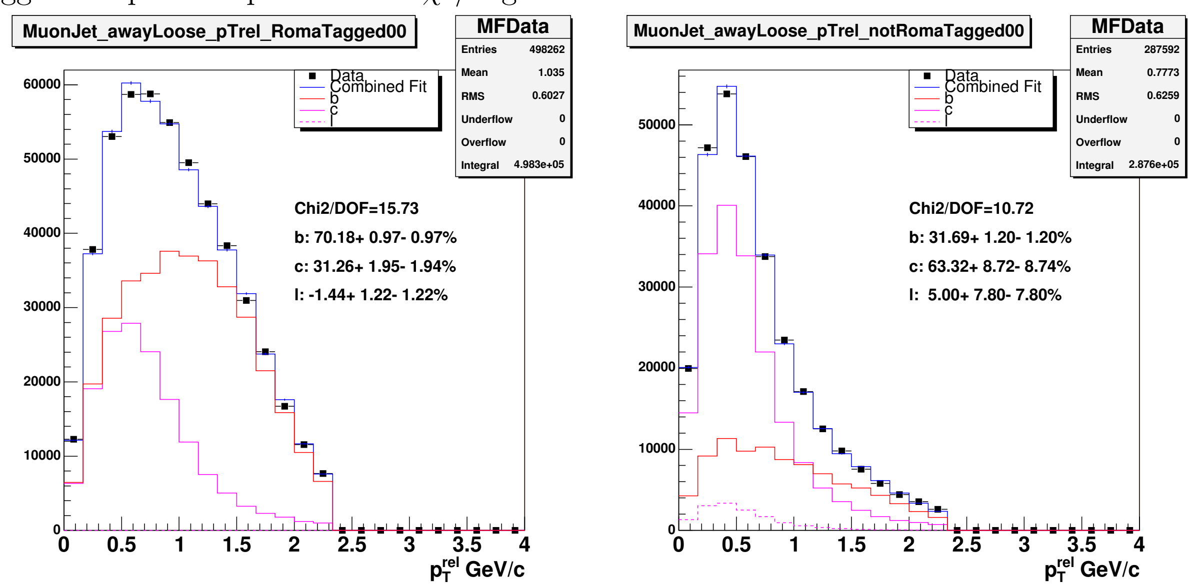
Table 20: Data and Monte Carlo calculations using definition of RomaNNout3>0.0 for tagged in the electron p_T^{rel} method.

8.1.2 Muon p_T^{rel}

Using the definition of RomaNN3out>0.0 for tagged, the data distribution and best fit results are shown in Fig. 53a for tagged jets and Fig. 53b for not-tagged jets. Table 21 summarizes the best fit results.

With 498,262 tagged jets in data, the best fit finds a b-fraction of $70.2 \pm 1.0\%$, a charm-fraction of $31.3 \pm 2.0\%$, and a light-fraction of $-1.4 \pm 1.2\%$ (consistent with zero), with 14 bins of 3/18 GeV/c and 2 fit parameters (f_b and f_c). The χ^2/degree-of-freedom is 15.7.

With 287,592 not-tagged jets in data, the best fit finds a b-fraction of $31.7 \pm 1.2\%$, a charm-fraction of $63.3 \pm 8.7\%$, and a light-fraction of $5.0 \pm 7.8\%$, with 14 bins of 3/18 GeV/c and 2 fit parameters (f_b and f_c). The not-tagged sample has a lower b-fraction than the tagged sample as expected. The χ^2/degree-of-freedom is 10.7.



(a) Tagged, total # of jets = 498262, b fraction = 70.2%

(b) Not Tagged, total # of jets = 287592, b fraction = 31.7%

Figure 53: Fitting b and non-b components to muon p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged.

The data efficiency for this sample of jets is then

$$\epsilon_{data} = \frac{0.702 \times 498262}{0.702 \times 498262 + 0.317 \times 287592} = \frac{349692}{349692 + 91129} = 0.793 \pm 0.007 \qquad (12)$$

In MC, the efficiency is estimated from the number of OBSP b-jets (passing the same kinematic cuts)

$$\epsilon_{MC} = \frac{35480}{35480 + 4440} = 0.889 \pm 0.002 \tag{13}$$

Therefore, the naive scale factor is

$$SF_{muon} = \frac{\epsilon_{data}}{\epsilon_{MC}} = \frac{0.795}{0.889} = 0.893 \pm 0.008 \tag{14}$$

Sample	*#tagged*	*%b*	*#b tagged*	*#not-tagged*	*%b*	*#b not-tagged*	ϵ_b
MC	35480	n/a	35480	4440	n/a	4440	0.889
data	498262	70.2	349692	287592	31.7	91129	0.793

Table 21: Data and Monte Carlo calculations using definition of RomaNNout3>0.0 for tagged in the muon p_T^{rel} method.

8.2 Other Operating Points

We have measured the MC efficiency, data efficiency, and scale factors for muon jets and hadronic jets, with the hadronic jets represented by electron jets. The scale factors are flat for a range of operating points from RomaNN3out>-0.30 to RomaNN3out>0.50, shown in Fig. 54, 55. This range can be seen to be useful from the RomaNN3out normalized distribution of MC b,c,l-jets in Fig. 3, where there is a good signal to background ratio without sacrificing too much signal.

It is convenient to express the scale factor for any RomaNNcut in terms of the scale factor at RomaNNcut=0.0. Table 22 lists the SF-Multipliers for cuts in a range of operating points from RomaNN3out>-0.30 to RomaNN3out>0.50.

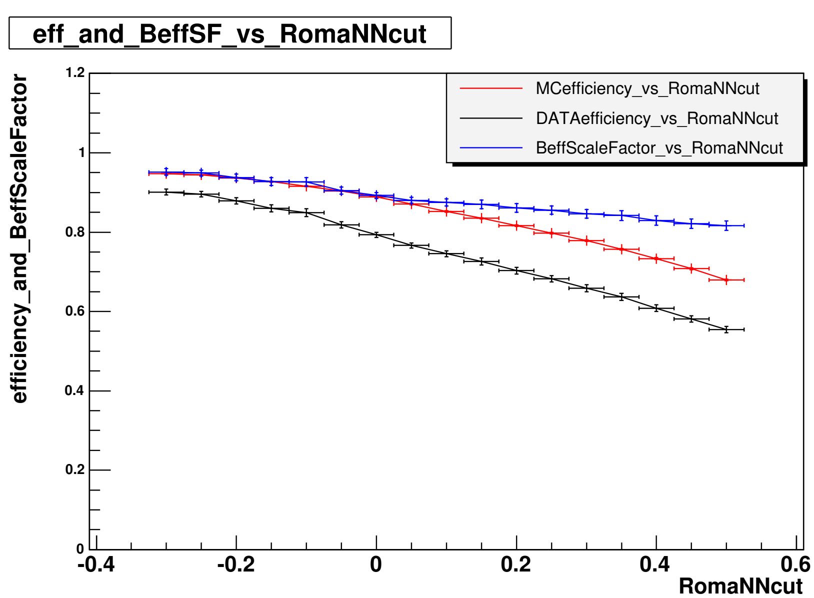


Figure 54: Scale Factor as a function of RomaNN3out cut value, muon p_T^{rel} method.

RomaNNcut	*Electron SF-Multiplier*	*Muon SF-Multiplier*
-0.30	1.042	1.066
-0.25	1.021	1.063
-0.20	1.020	1.051
-0.15	1.013	1.040
-0.10	1.006	1.038
-0.05	1.000	1.014
0.00	1.000	1.000
0.05	0.987	0.986
0.10	1.005	0.980
0.15	1.012	0.975
0.20	1.014	0.965
0.25	1.004	0.958
0.30	0.996	0.948
0.35	0.988	0.943
0.40	0.979	0.929
0.45	0.965	0.920
0.50	0.962	0.915

Usage: $SF(RomaNNcut) = SF(RomaNNcut = 0.0) \times SF\text{–}Multiplier(RomaNNcut)$

Table 22: SF-Multipliers for cuts in a range of operating points from RomaNN3out>-0.30 to RomaNN3out>0.50.

8.3 Additional Cross checks

8.3.1 Tight SecVtx Scale Factor Measurement using Electron p_T^{rel}

Using the electron p_T^{rel} method as was done for the RomaNN, the MC and data efficiency for tagging b-jets using Tight SecVtx are calculated according to equation (4), the best fit results then produce a central scale factor of 0.928±0.015(stat)±0.80(syst) (Fig. 56).

There is an issue that the electron p_T^{rel} spectrum depends on the E_T distribution of the sample, therefore the fits in the individual E_T bins are more accurate because they are not affected by the difference in E_T distribution between data and MC templates.

Fig. 57 shows a trend of the scale factor as a function of jet E_T. Fitting a flat line we get a scale factor of 0.855±0.013(stat)±0.069(syst), in agreement with the muon p_T^{rel} blessed results 0.932±0.022(stat)±0.049(syst) [2]. Furthermore, the scale factor does not depends on jet E_T, as shown in Fig. 57.

Sample	*#tagged*	*%b*	*#b tagged*	*#not-tagged*	*%b*	*#b not-tagged*	ϵ_b
MC	13917	n/a	13917	30345	n/a	30345	0.314
data	47531	88.97	42288	199935	51.32	102607	0.292

Table 23: Data and Monte Carlo calculations using definition of SecvTag=1 for tagged, and SecvTag≠1 for not-tagged.

$$SF_{elec} = \frac{\epsilon_{data}}{\epsilon_{MC}} = \frac{0.292}{0.314} = 0.928 \pm 0.0153(stat) \tag{15}$$

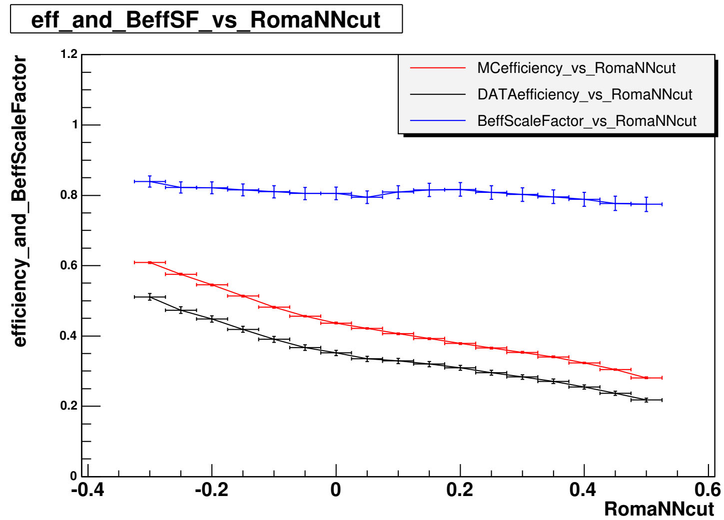
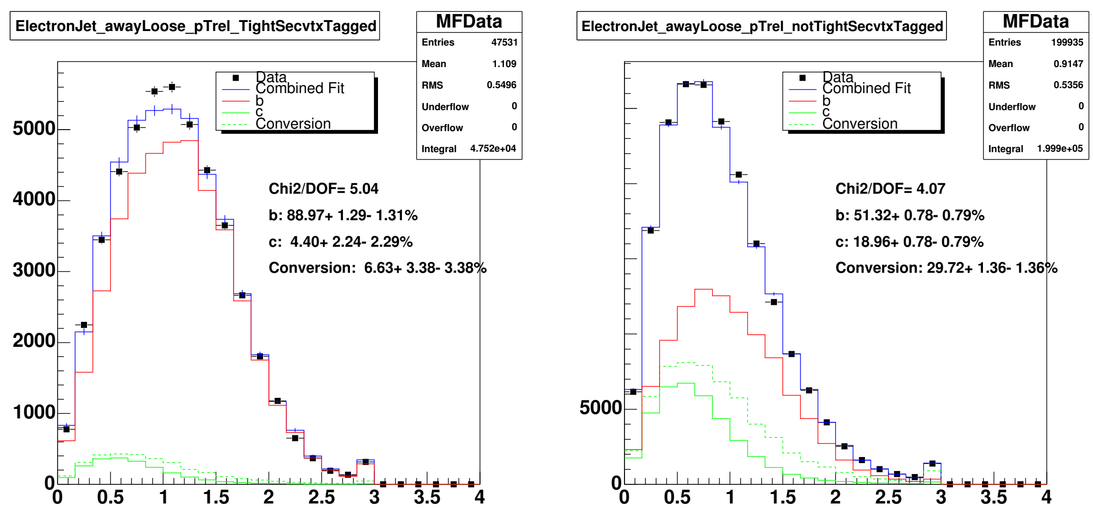


Figure 55: Scale Factor as a function of RomaNN3out cut value, using electron p_T^{rel}.



(a) Tagged, total # of jets = 47531, b fraction = 88.97%

(b) Not Tagged, total # of jets = 199935, b fraction = 51.32%

Figure 56: Fitting b and non-b components to elec p_T^{rel} data, using definition of SecvTag=1 for tagged, and SecvTag≠1 for not-tagged.

8.3.2 Jet Direction Measurement

In addition to the momentum being correct, an accurate p_T^{rel} depends on the jet direction being correct as well. In this study, the jet direction is measured using the calorimeter. The previous SecVtx scale factor measurements [2], [3] using the muon p_T^{rel} technique studied an alternative method of determining the jet direction: it can be measured from the tracks lying inside the jet cone.

However, in the electron p_T^{rel} technique, this alternative method of determining the jet direction is not possible. This is due to the fact that bremsstrahlung photons are a significant part of an electron-jet, and the calorimeter is needed to capture the photon information to accurately determine the jet direction. This is evident in the comparison of the p_T^{rel} templates (Fig. 58, 59, 60) where the electron in the track based reconstruction is much closer to the jet axis.

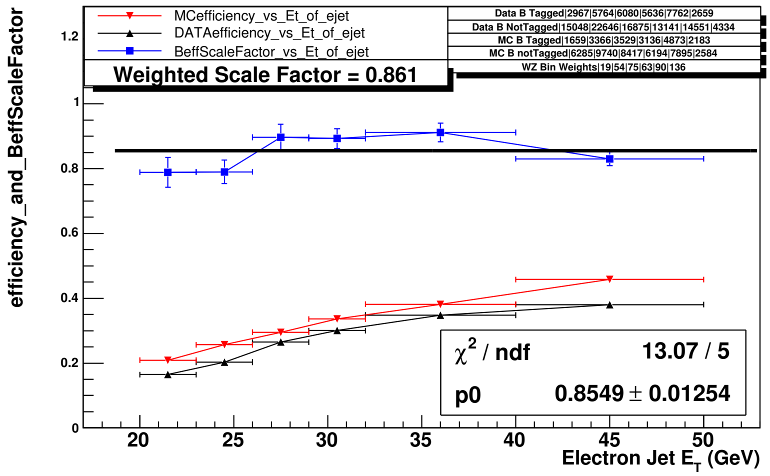
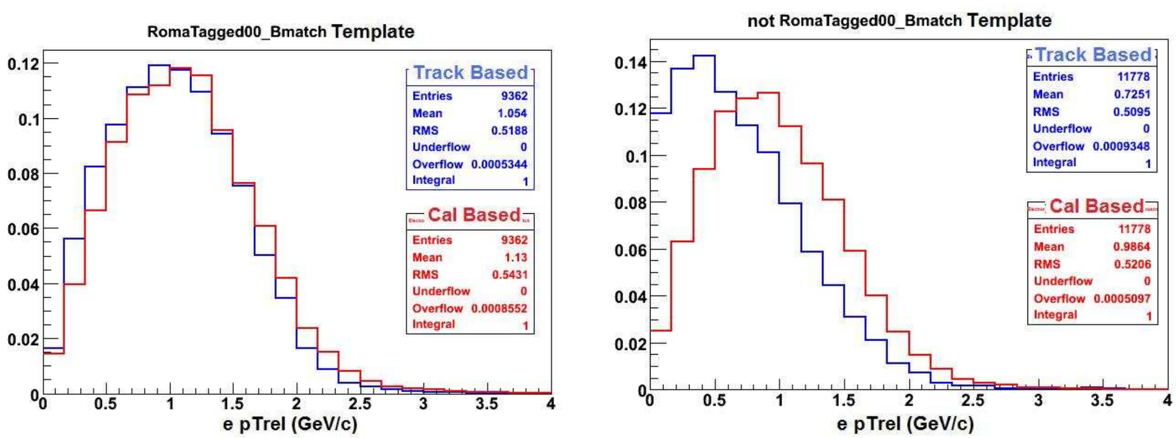


Figure 57: Scale Factor vs. transverse energy, which when weighted with the E_T spectrum of the b-jets in WZ MC is 0.861. The Scale Factor is flat against transverse energy.



(a) RomaNN3out>0.00 (tagged) b-template comparison.

(b) RomaNN3out$\leq$0.00 (not-tagged) b-template comparison.

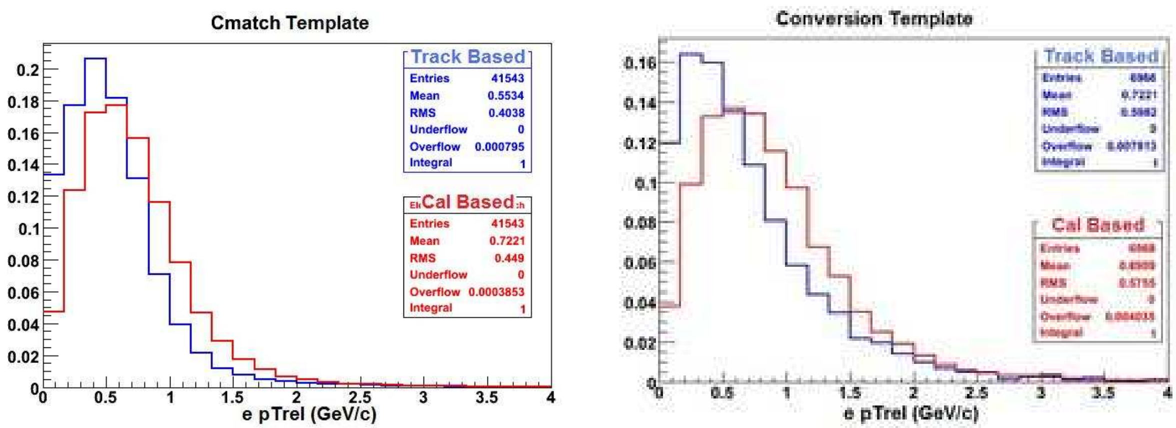
Figure 58: A comparison of the calorimeter and track based p_T^{rel} b-templates.

8.3.3 Alternate Electron p_T^{rel} Observable

We checked the robustness of our p_T^{rel} results by considering an alternate observable, p_T^{rel}/E_T, which does not depend on the jet E_T as much Fig. 61. (This was done using data up to p17). Using this alternate observable, we calculated the central value of the scale factor to be 0.812 ± 0.018(stat), and plotted Fig. 63 shows a trend of the scale factor as a function of jet E_T. Fitting a flat line we get a scale factor of 0.803 ± 0.009(stat), both are consistent when comparing it to scale factor of 0.793 ± 0.018(stat) from Fig. 16.

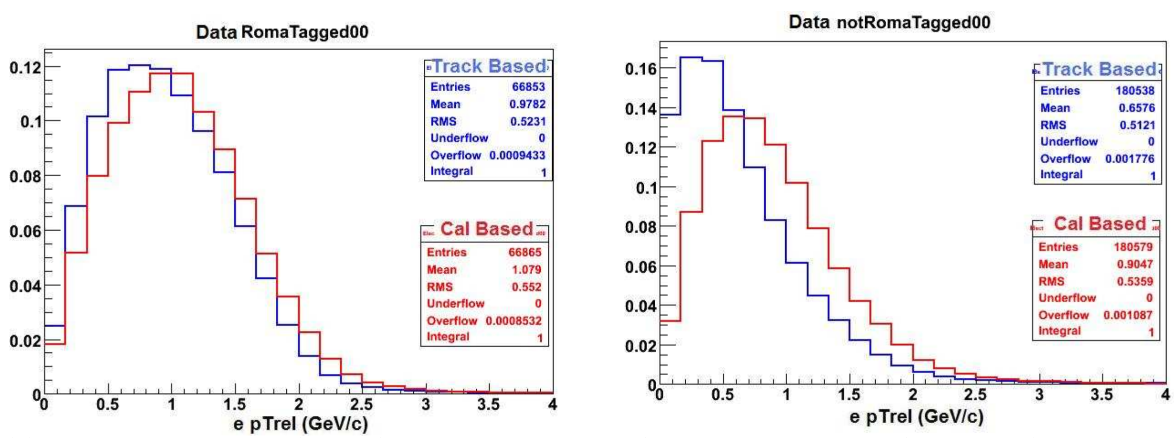
Also we show in Fig. 64 a trend of the scale factor as a function of the number of tracks in the jet.

$$SF_{elec} = \frac{\epsilon_{data}}{\epsilon_{MC}} = \frac{0.356}{0.437} = 0.812 \pm 0.018(stat) \tag{16}$$



(a) c-template comparison. (b) l-template comparison.

Figure 59: A comparison of the calorimeter and track based p_T^{rel} c/l-templates.

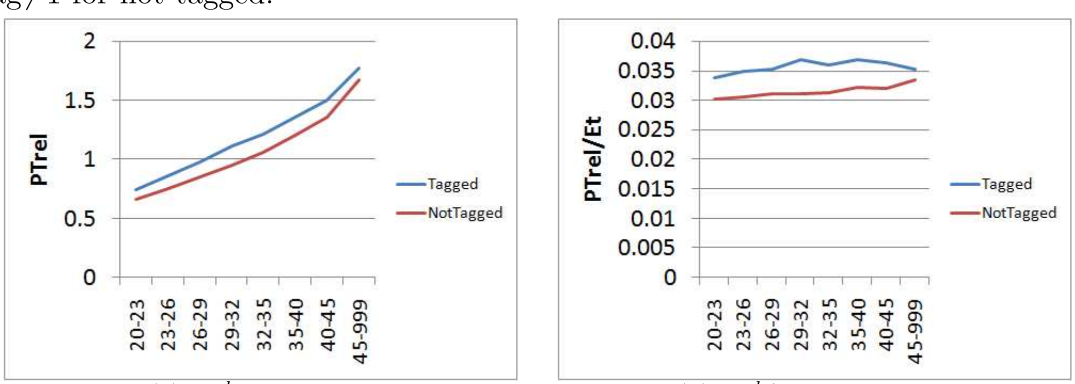


(a) RomaNN3out>0.00 (tagged) data distribution comparison. (b) RomaNN3out≤0.00 (not-tagged) data distribution comparison.

Figure 60: A comparison of the calorimeter and track based p_T^{rel} distribution in data.

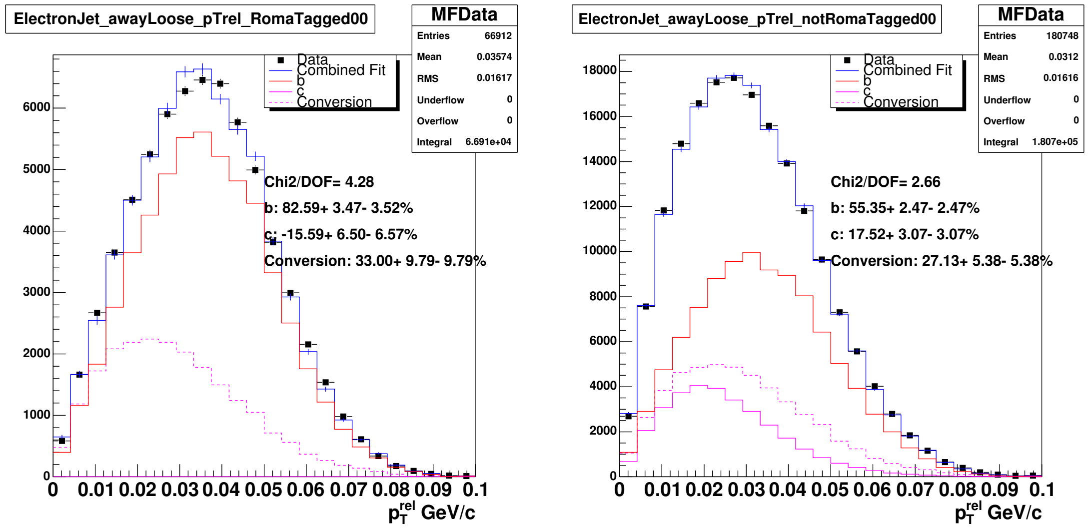
Sample	*#tagged*	*%b*	*#b tagged*	*#not-tagged*	*%b*	*#b not-tagged*	ϵ_b
MC	19330	n/a	19330	24932	n/a	24932	0.437
data	66912	82.6	55262	180748	55.4	100044	0.356

Table 24: Data and Monte Carlo calculations using definition of SecvTag=1 for tagged, and SecvTag≠1 for not-tagged.



(a) p_T^{rel} vs Jet E_T. (b) p_T^{rel}/E_T vs Jet E_T.

Figure 61: A comparison of the p_T^{rel} and p_T^{rel}/E_T vs. Jet E_T.



(a) Tagged, total # of jets = 66912, b fraction = 82.6%

(b) Not Tagged, total # of jets = 180748, b fraction = 55.4%

Figure 62: Fitting b and non-b components to elec p_T^{rel} data, using RomaNNCut=0.0.

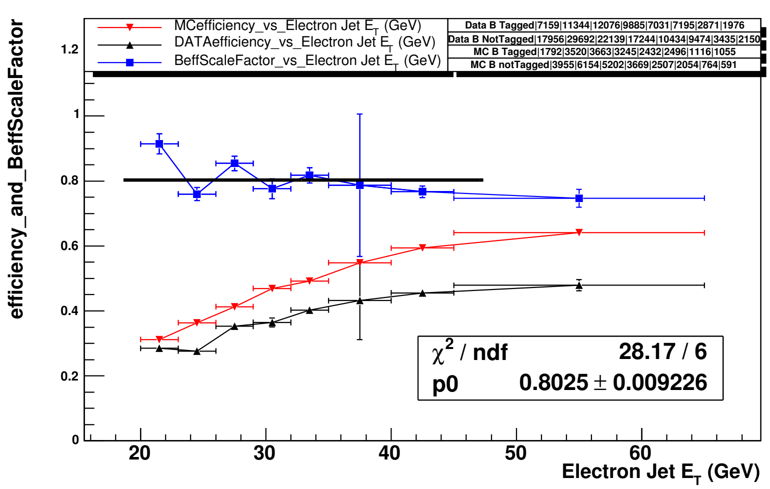


Figure 63: Scale Factor vs. transverse energy, using electron p_T^{rel}/E_T.

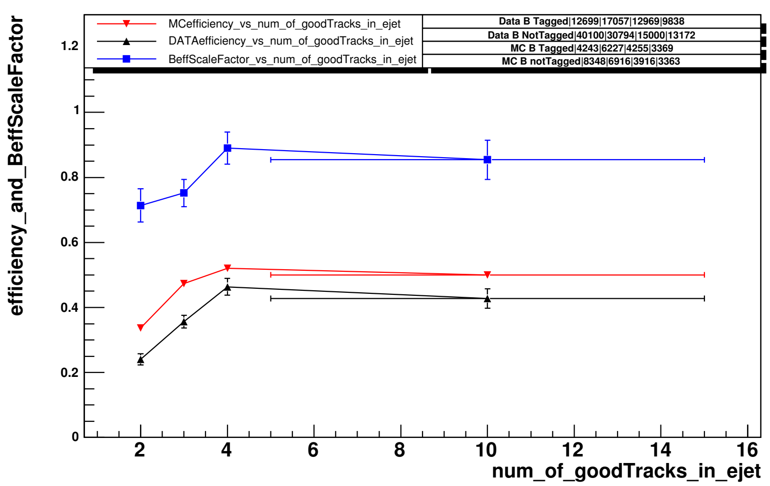
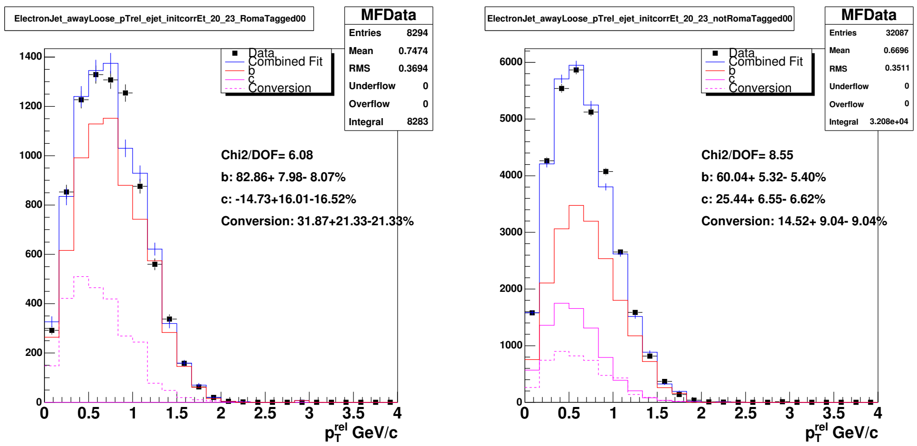


Figure 64: Scale Factor vs. Number of RomaNN good tracks, using electron p_T^{rel}/E_T.

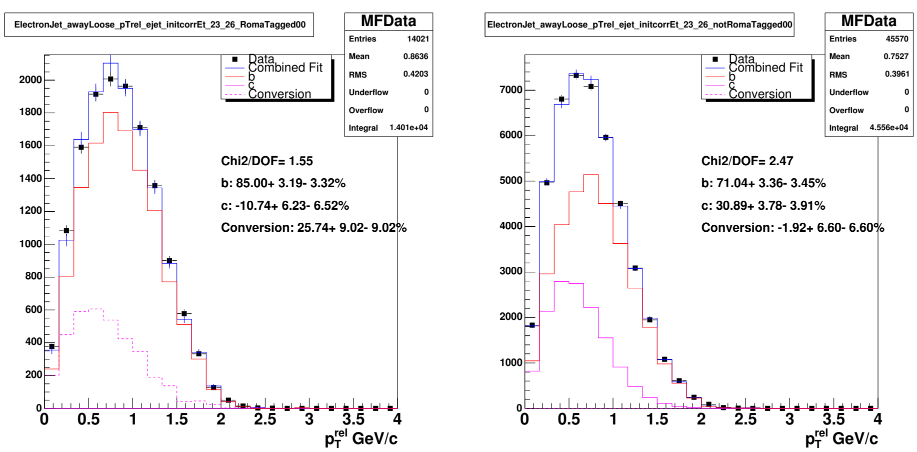
8.4 Supporting plots

8.4.1 Fits From Electron p_T^{rel} Trends



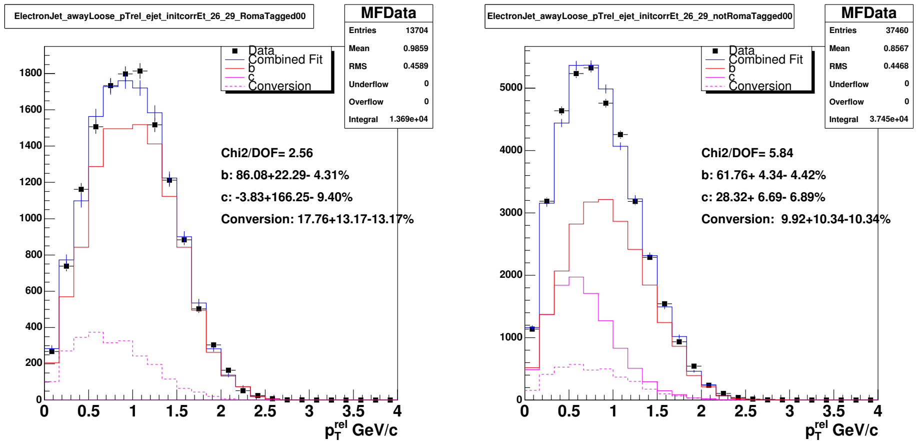
(a) Tagged (b) Not Tagged

Figure 65: Fitting b and non-b components to elec p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of $20 \leq E_T < 23$.



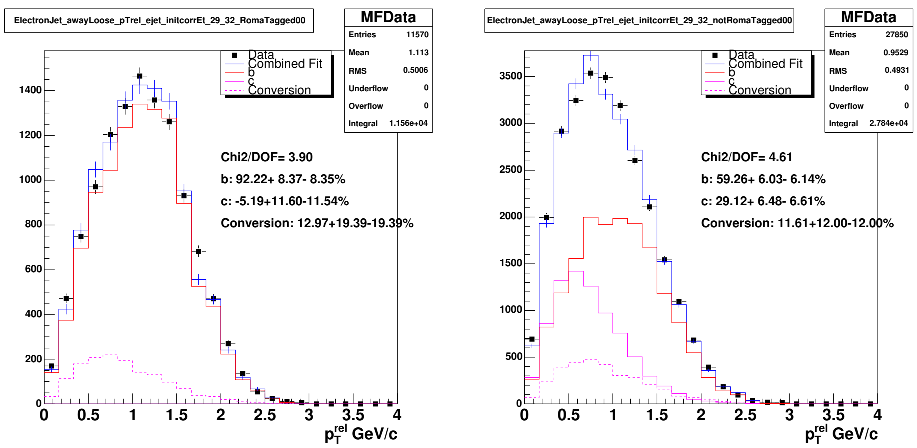
(a) Tagged (b) Not Tagged

Figure 66: Fitting b and non-b components to elec p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of $23 \leq E_T < 26$.



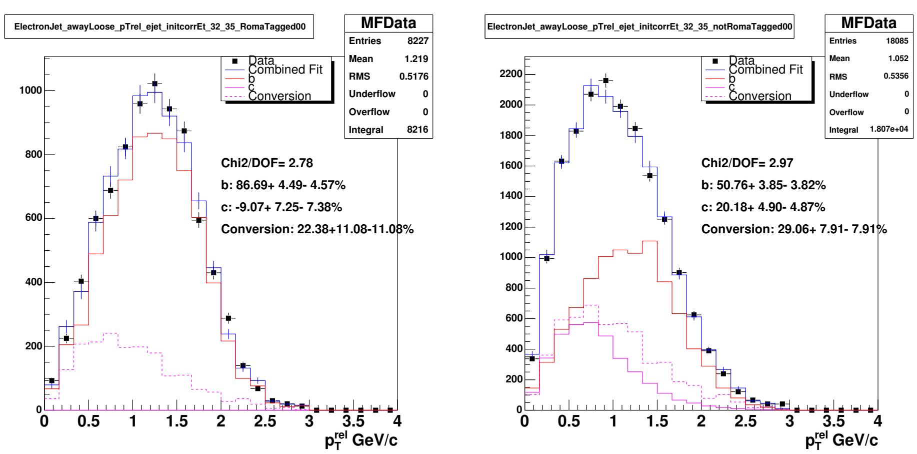
(a) Tagged (b) Not Tagged

Figure 67: Fitting b and non-b components to elec p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of $26 \leq E_T < 29$.



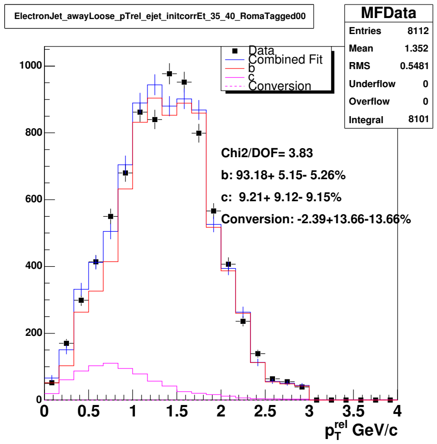
(a) Tagged (b) Not Tagged

Figure 68: Fitting b and non-b components to elec p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of $29 \leq E_T < 32$.

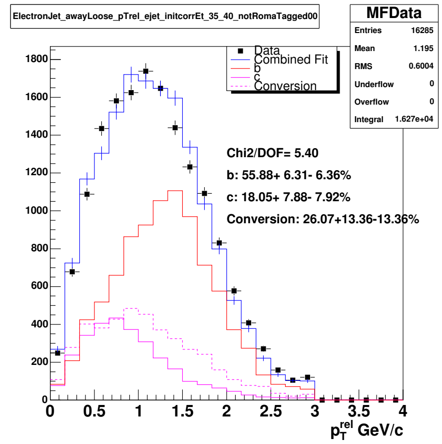


(a) Tagged (b) Not Tagged

Figure 69: Fitting b and non-b components to elec p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of $32 \leq E_T < 35$.

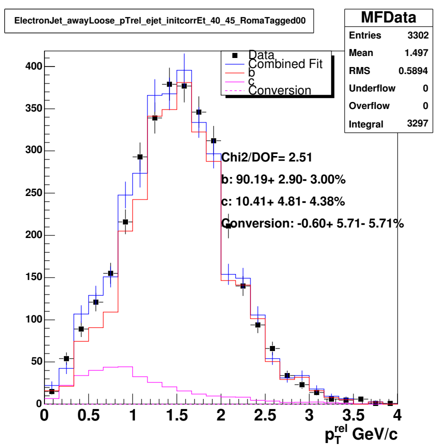


(a) Tagged

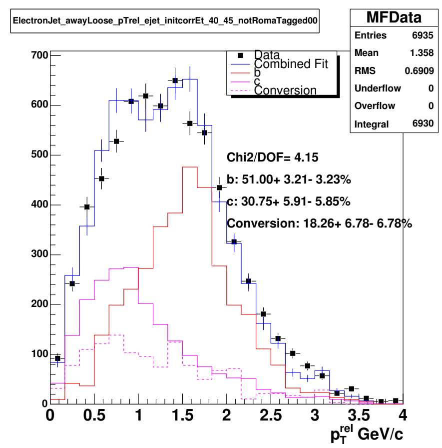


(b) Not Tagged

Figure 70: Fitting b and non-b components to elec p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of $35 \leq E_T < 40$.

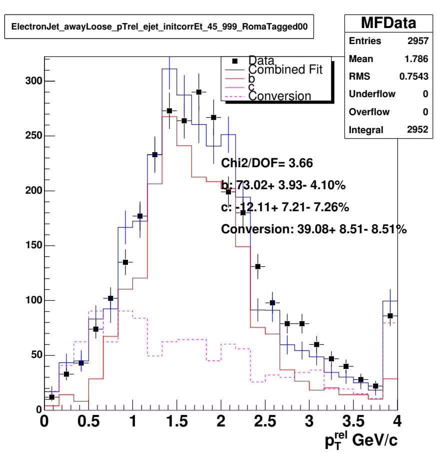


(a) Tagged

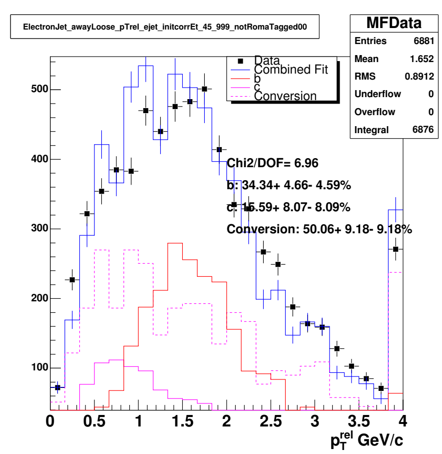


(b) Not Tagged

Figure 71: Fitting b and non-b components to elec p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of $40 \leq E_T < 45$.

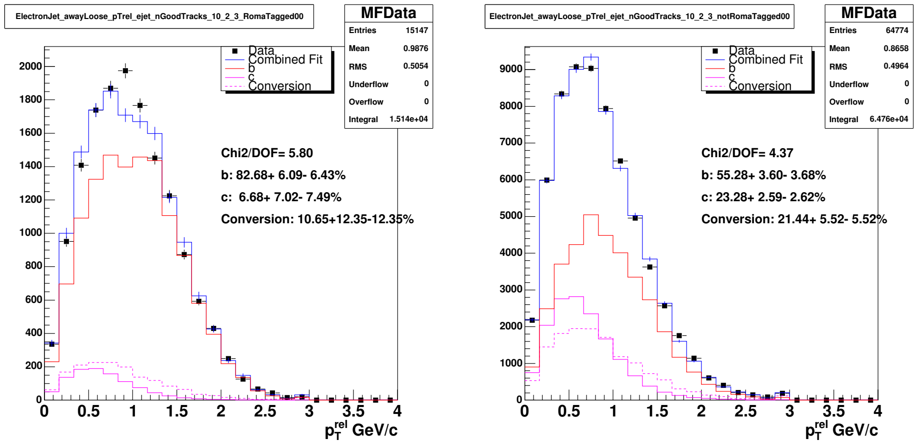


(a) Tagged



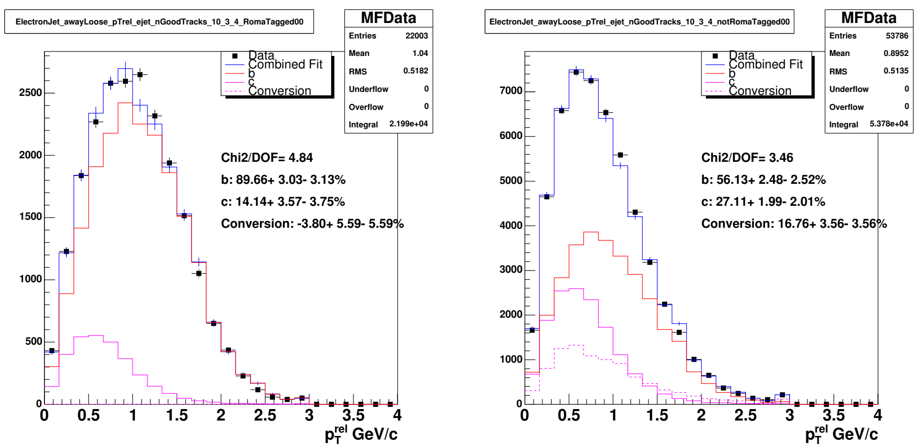
(b) Not Tagged

Figure 72: Fitting b and non-b components to elec p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of $E_T \geq 45$.



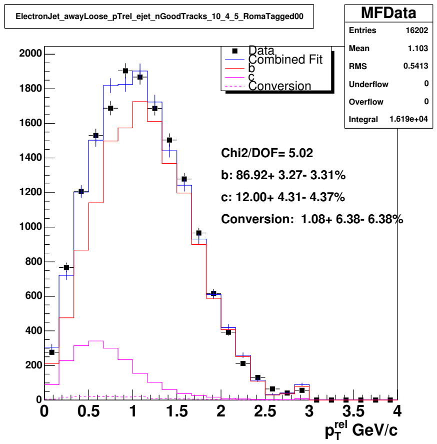
(a) Tagged (b) Not Tagged

Figure 73: Fitting b and non-b components to elec p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of nGoodTracks = 2.

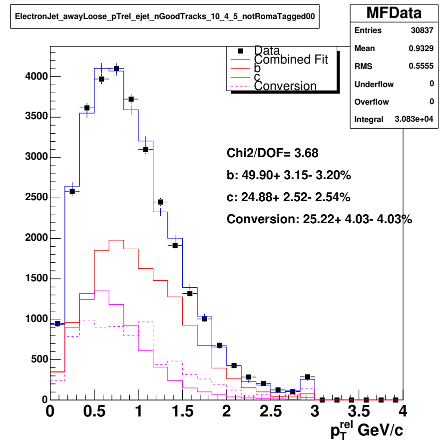


(a) Tagged (b) Not Tagged

Figure 74: Fitting b and non-b components to elec p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of nGoodTracks = 3.

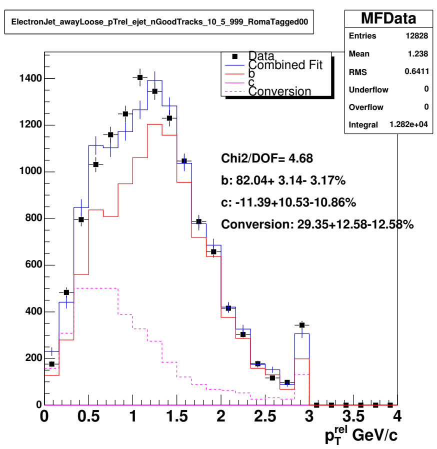


(a) Tagged

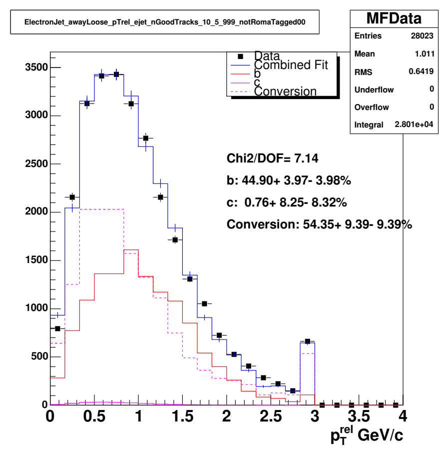


(b) Not Tagged

Figure 75: Fitting b and non-b components to elec p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of nGoodTracks = 4.

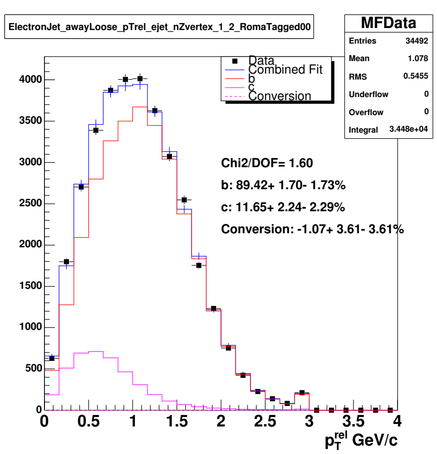


(a) Tagged

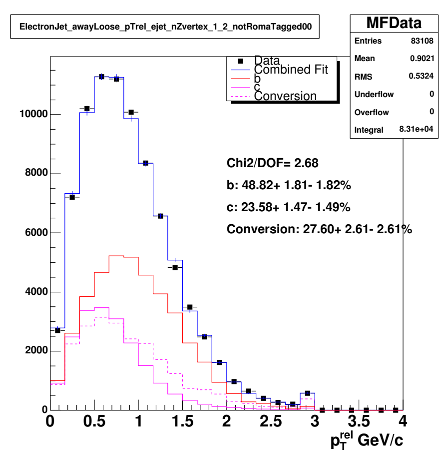


(b) Not Tagged

Figure 76: Fitting b and non-b components to elec p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of nGoodTracks $\geq$ 5.

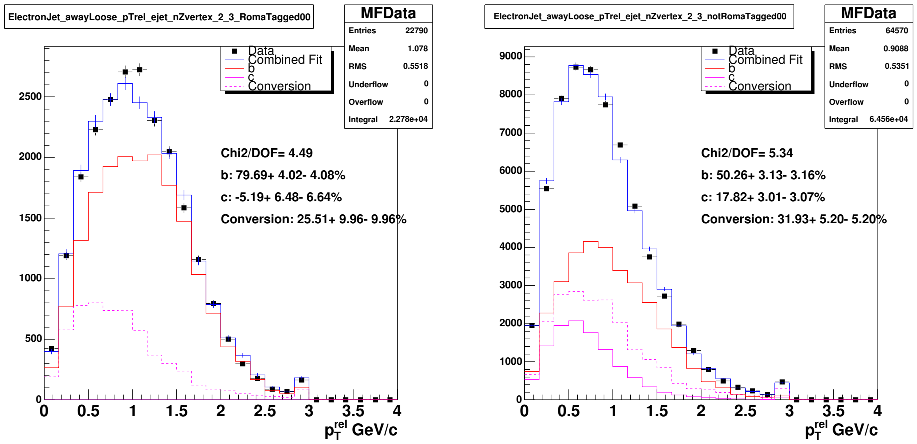


(a) Tagged



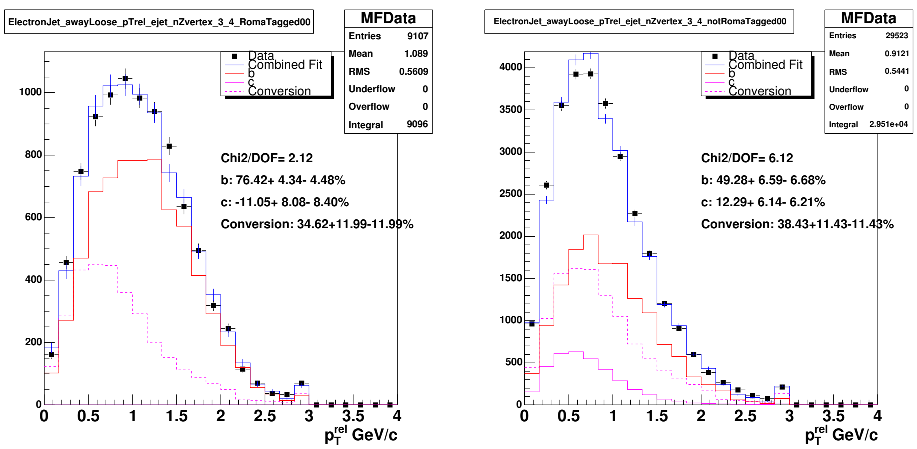
(b) Not Tagged

Figure 77: Fitting b and non-b components to elec p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of nZvertex = 1.



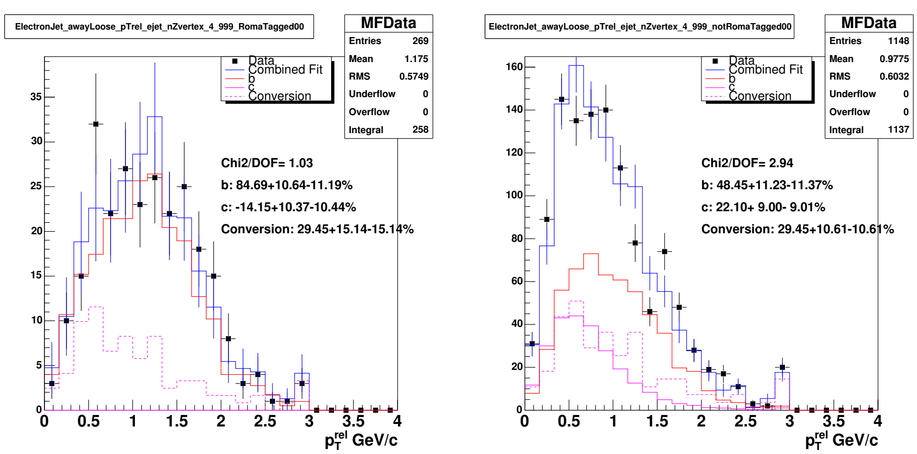
(a) Tagged (b) Not Tagged

Figure 78: Fitting b and non-b components to elec p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of nZvertex = 2.



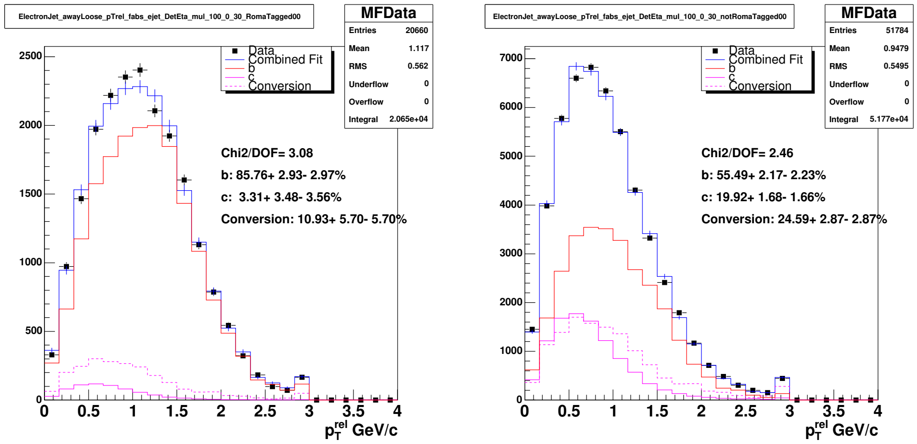
(a) Tagged (b) Not Tagged

Figure 79: Fitting b and non-b components to elec p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of nZvertex = 3.



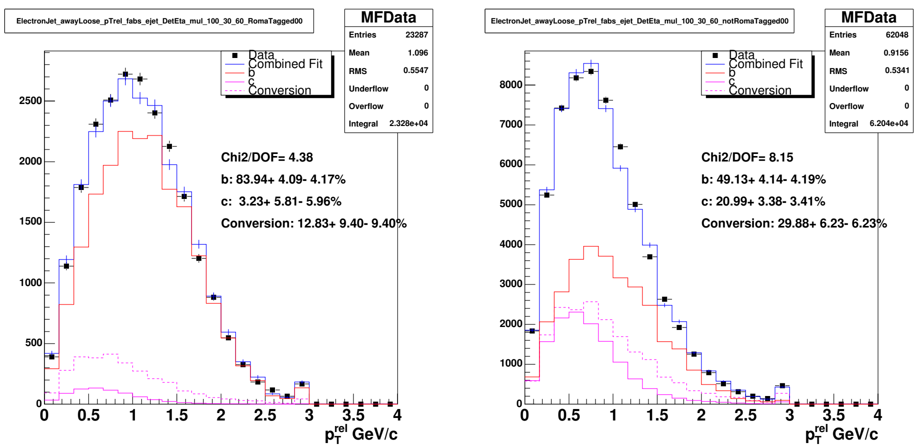
(a) Tagged (b) Not Tagged

Figure 80: Fitting b and non-b components to elec p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of nZvertex $\geq$ 4.



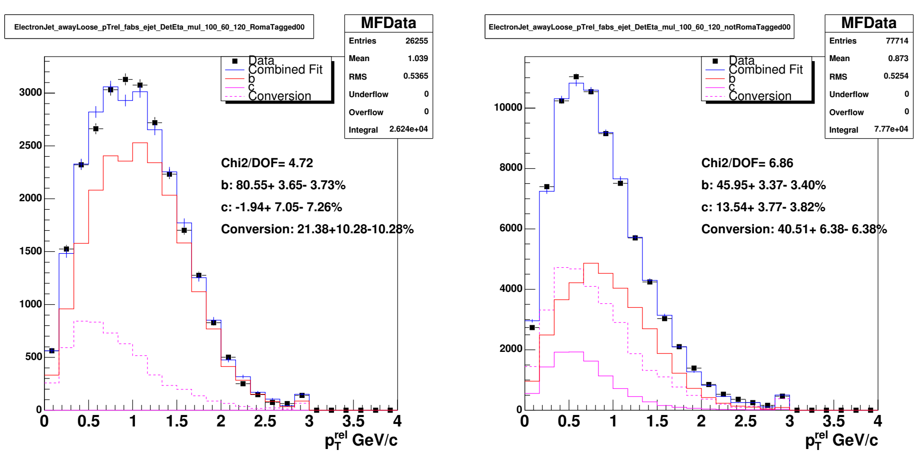
(a) Tagged (b) Not Tagged

Figure 81: Fitting b and non-b components to elec p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of $0 \leq \eta < 0.30$.



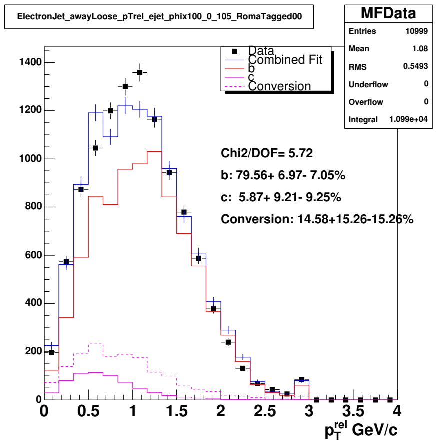
(a) Tagged (b) Not Tagged

Figure 82: Fitting b and non-b components to elec p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of $0.30 \leq \eta < 0.60$.

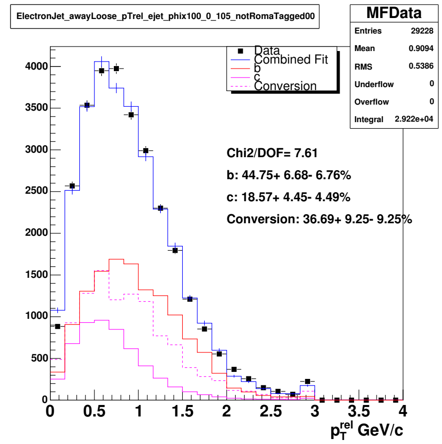


(a) Tagged (b) Not Tagged

Figure 83: Fitting b and non-b components to elec p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of $0.60 \leq \eta < 1.20$.

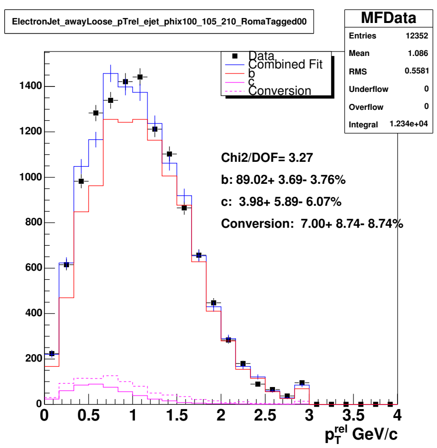


(a) Tagged

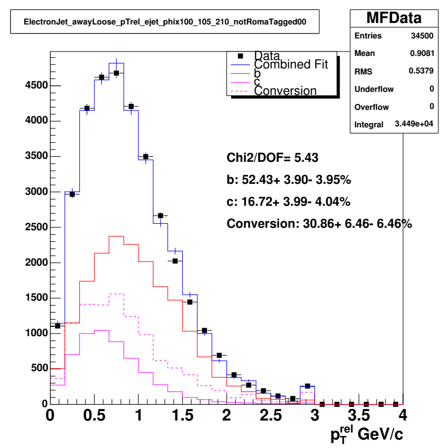


(b) Not Tagged

Figure 84: Fitting b and non-b components to elec p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of $0 \leq \phi < \frac{\pi}{3}$.

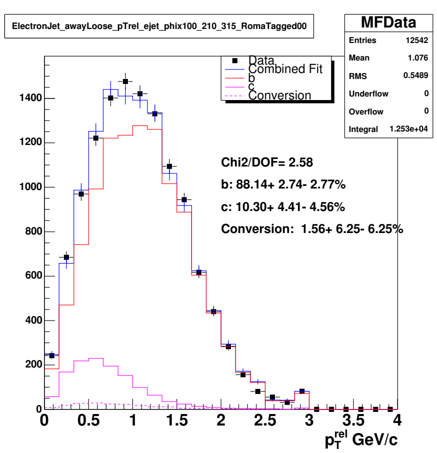


(a) Tagged

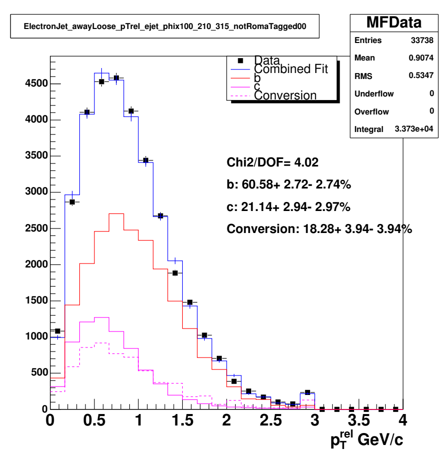


(b) Not Tagged

Figure 85: Fitting b and non-b components to elec p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of $\frac{\pi}{3} \leq \phi < \frac{2\pi}{3}$.

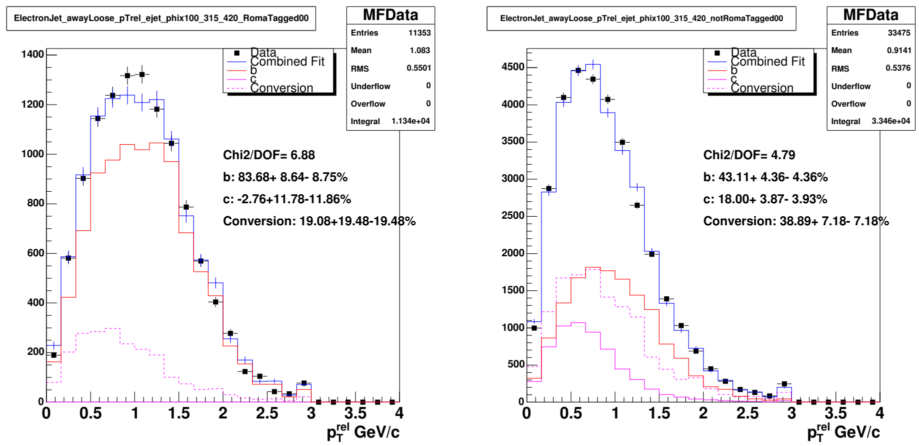


(a) Tagged



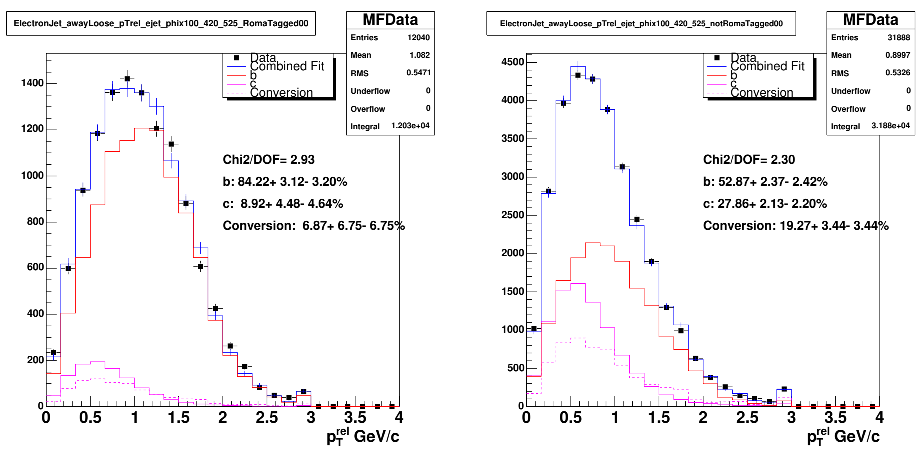
(b) Not Tagged

Figure 86: Fitting b and non-b components to elec p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of $\frac{2\pi}{3} \leq \phi < \pi$.



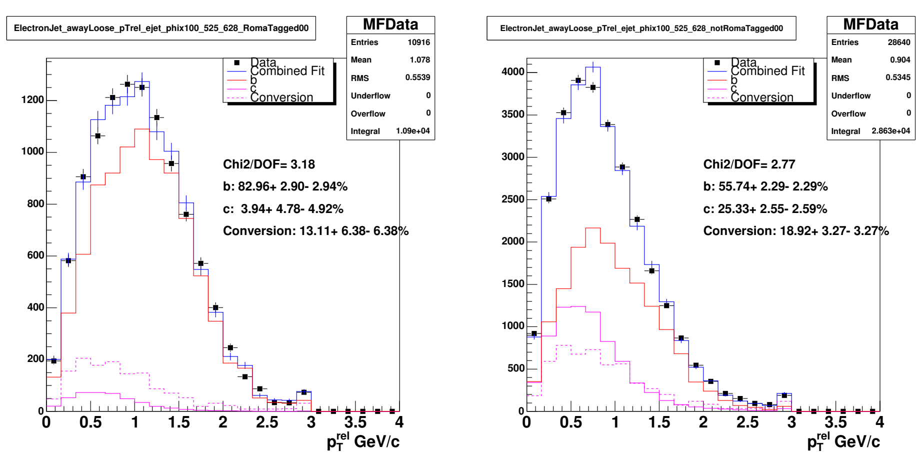
(a) Tagged (b) Not Tagged

Figure 87: Fitting b and non-b components to elec p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of $\pi \leq \phi < \frac{4\pi}{3}$.



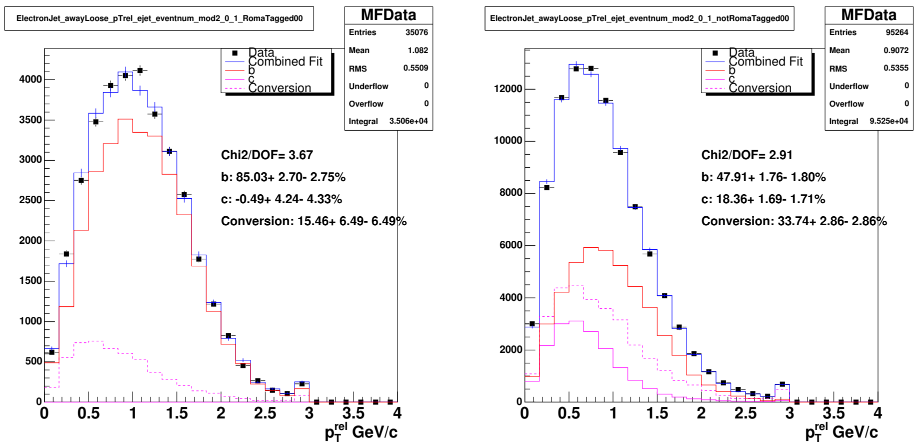
(a) Tagged (b) Not Tagged

Figure 88: Fitting b and non-b components to elec p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of $\frac{4\pi}{3} \leq \phi < \frac{5\pi}{3}$.



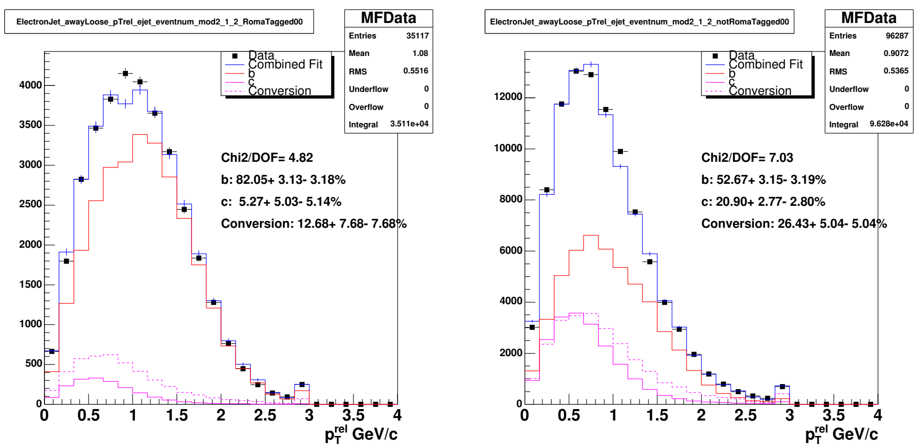
(a) Tagged (b) Not Tagged

Figure 89: Fitting b and non-b components to elec p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of $\frac{5\pi}{3} \leq \phi < 2\pi$.



(a) Tagged (b) Not Tagged

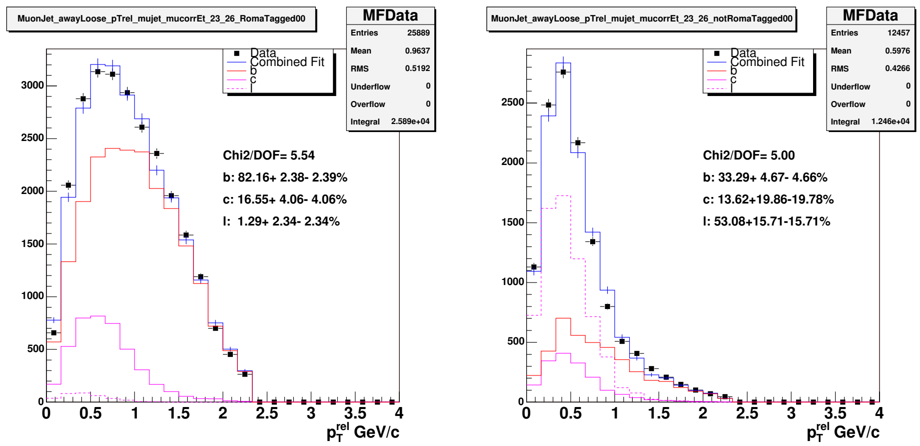
Figure 90: Fitting b and non-b components to elec p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Even event number bin.



(a) Tagged (b) Not Tagged

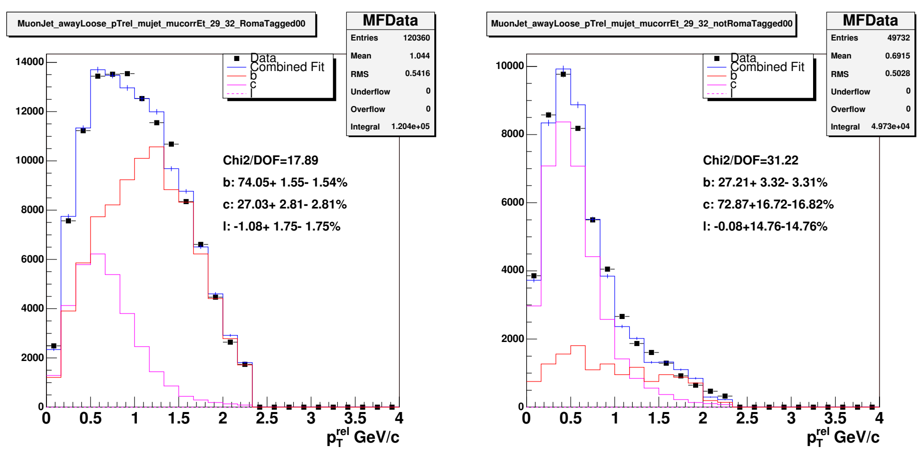
Figure 91: Fitting b and non-b components to elec p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Odd event number bin.

8.4.2 Fits From Muon p_T^{rel} Trends



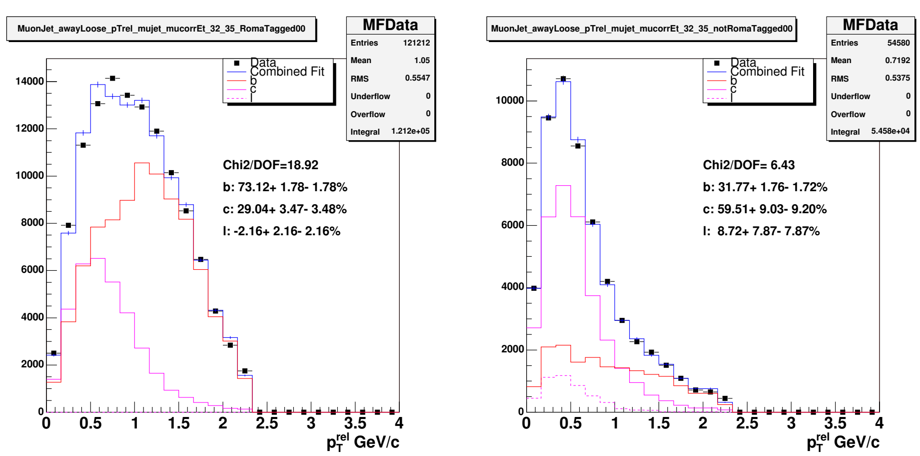
(a) Tagged (b) Not Tagged

Figure 92: Fitting b and non-b components to muon p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of $20 \leq E_T < 29$.



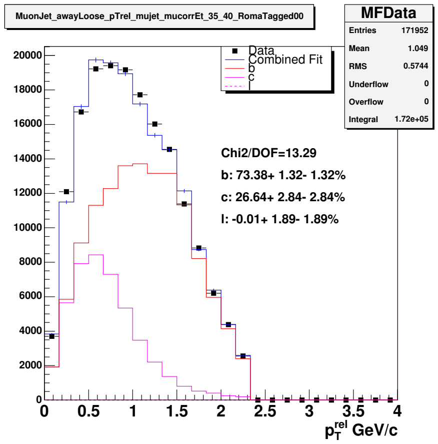
(a) Tagged (b) Not Tagged

Figure 93: Fitting b and non-b components to muon p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of $29 \leq E_T < 32$.

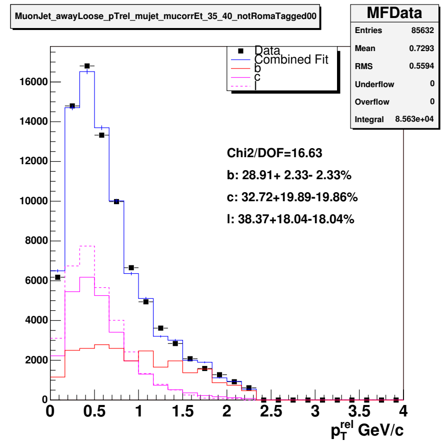


(a) Tagged (b) Not Tagged

Figure 94: Fitting b and non-b components to muon p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of $32 \leq E_T < 35$.

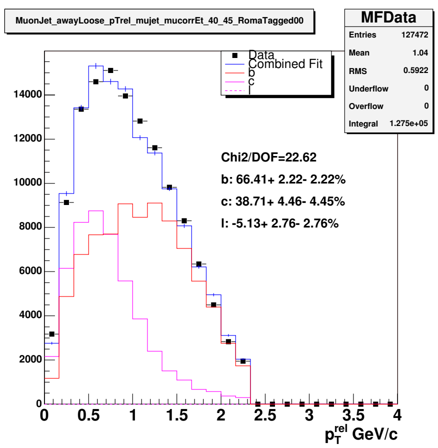


(a) Tagged

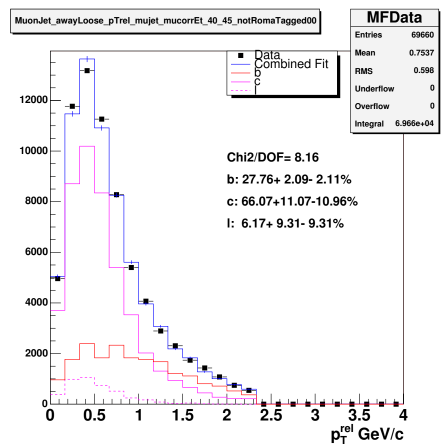


(b) Not Tagged

Figure 95: Fitting b and non-b components to muon p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of $35 \leq E_T < 40$.

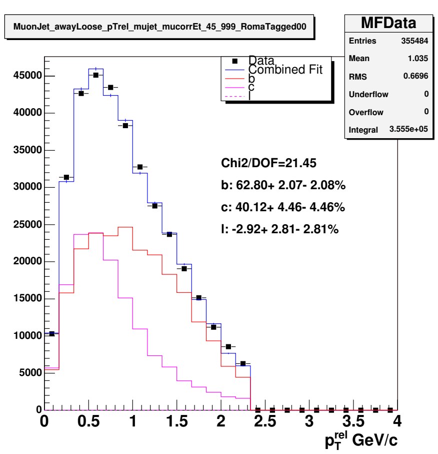


(a) Tagged

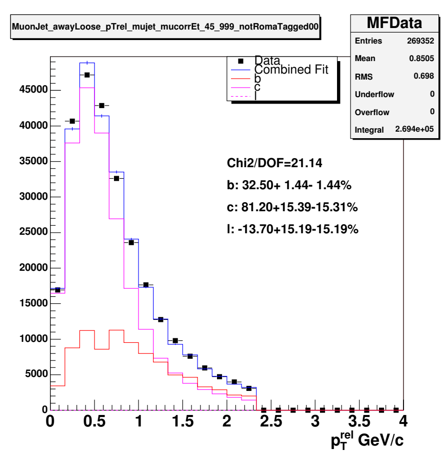


(b) Not Tagged

Figure 96: Fitting b and non-b components to muon p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of $40 \leq E_T < 45$.

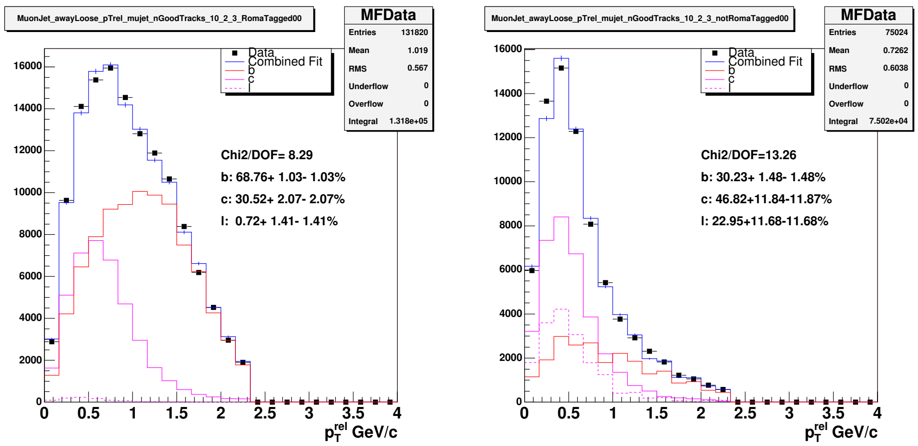


(a) Tagged



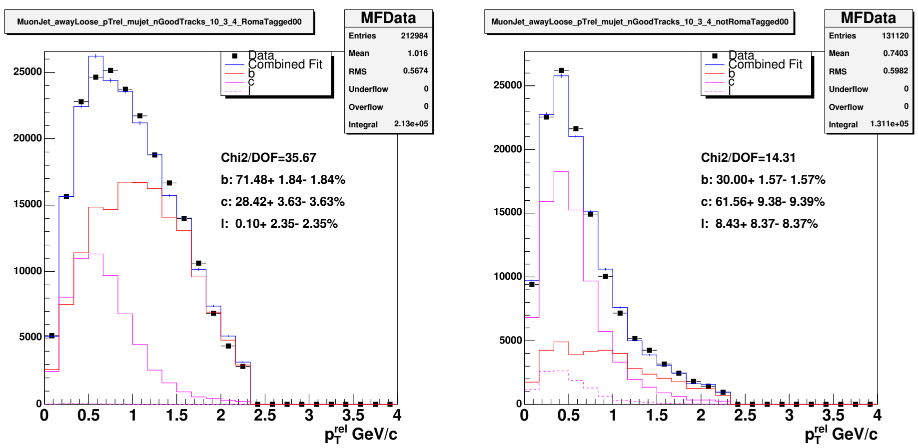
(b) Not Tagged

Figure 97: Fitting b and non-b components to muon p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of $E_T \geq 45$.



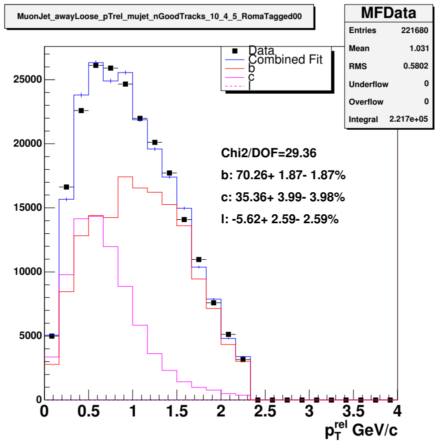
(a) Tagged (b) Not Tagged

Figure 98: Fitting b and non-b components to muon p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of nGoodTracks = 2.

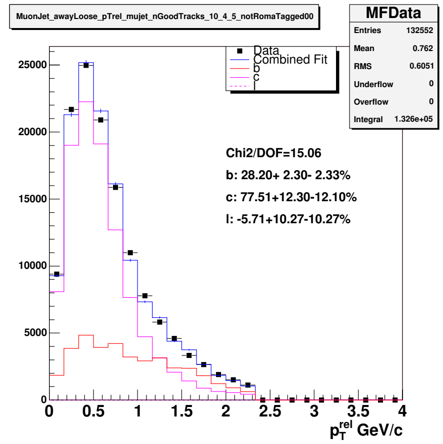


(a) Tagged (b) Not Tagged

Figure 99: Fitting b and non-b components to muon p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of nGoodTracks = 3.

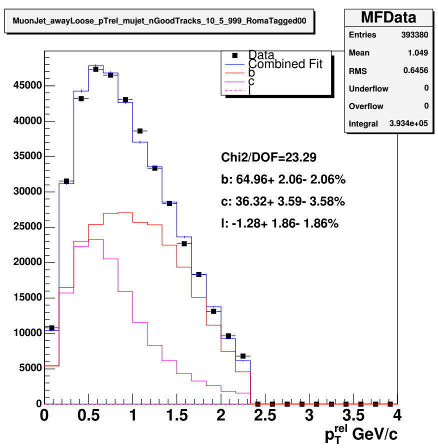


(a) Tagged

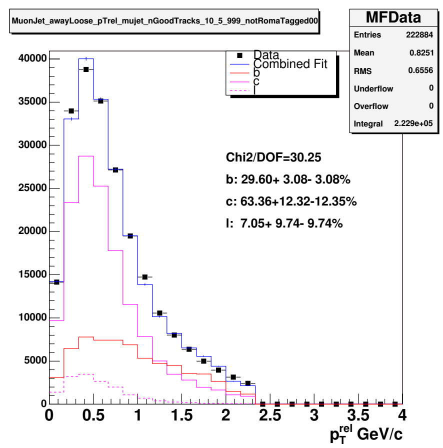


(b) Not Tagged

Figure 100: Fitting b and non-b components to muon p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of nGoodTracks = 4.

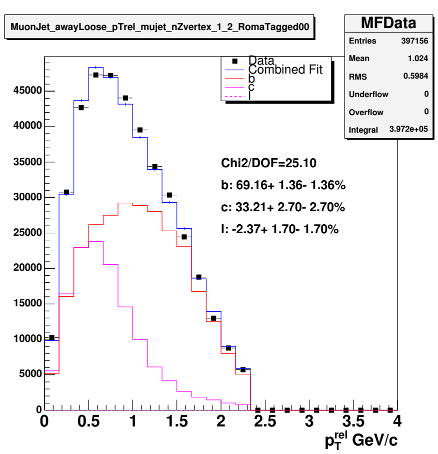


(a) Tagged

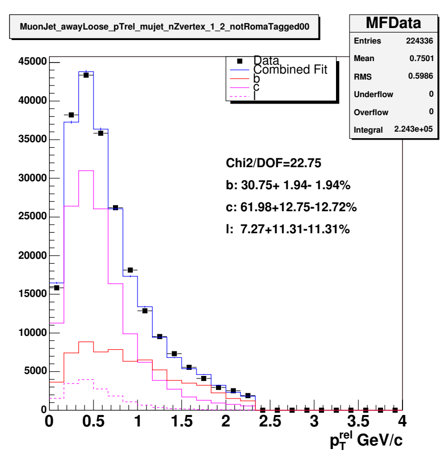


(b) Not Tagged

Figure 101: Fitting b and non-b components to muon p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of nGoodTracks $\geq$ 5.

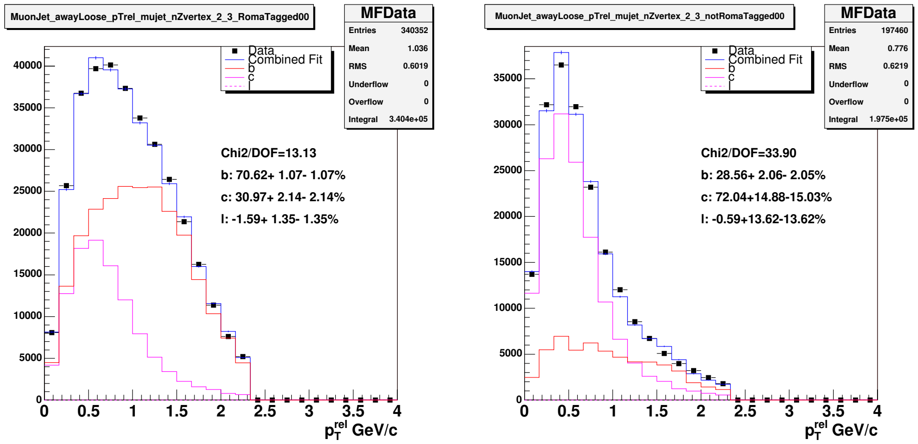


(a) Tagged



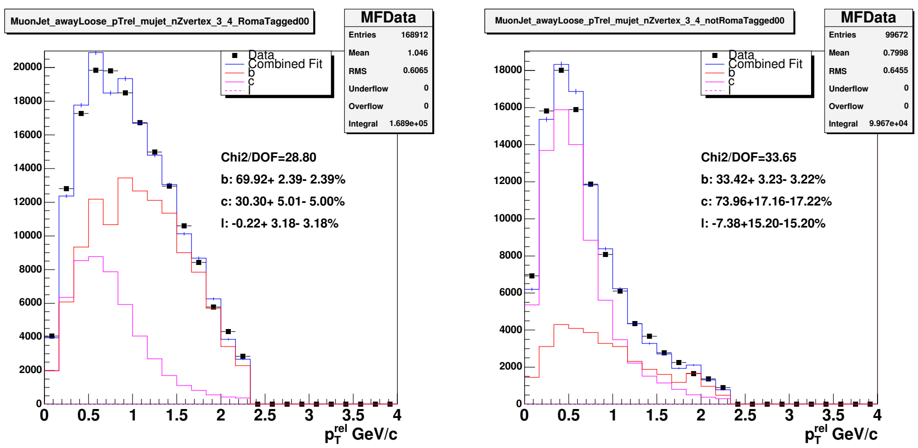
(b) Not Tagged

Figure 102: Fitting b and non-b components to muon p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of nZvertex = 1.



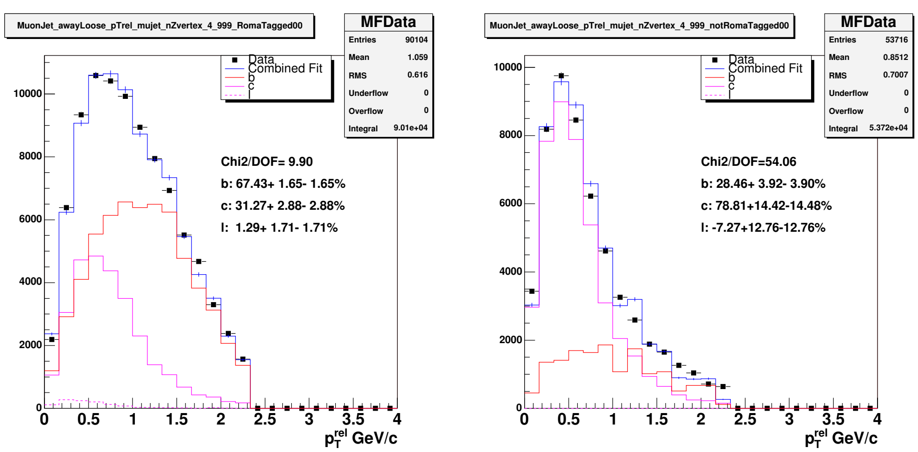
(a) Tagged (b) Not Tagged

Figure 103: Fitting b and non-b components to muon p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of nZvertex = 2.



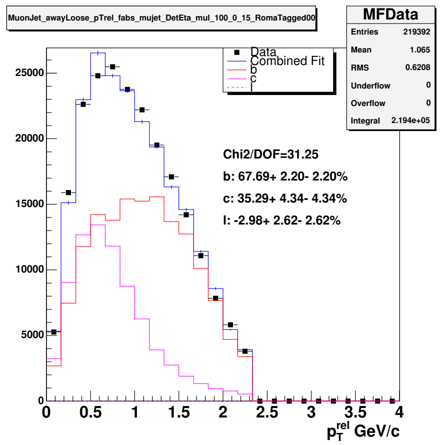
(a) Tagged (b) Not Tagged

Figure 104: Fitting b and non-b components to muon p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of nZvertex = 3.

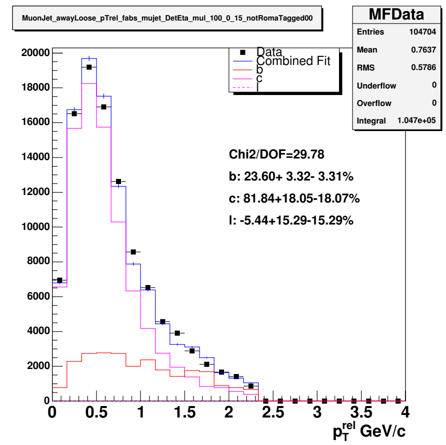


(a) Tagged (b) Not Tagged

Figure 105: Fitting b and non-b components to muon p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of nZvertex $\geq$ 4.

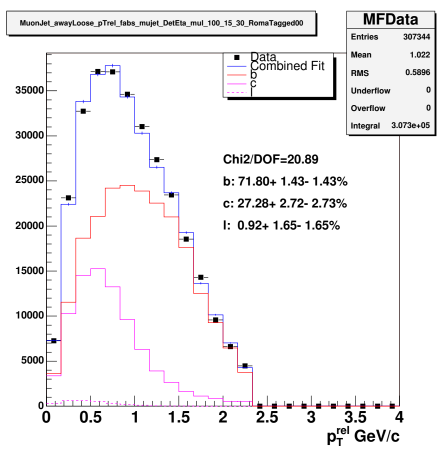


(a) Tagged

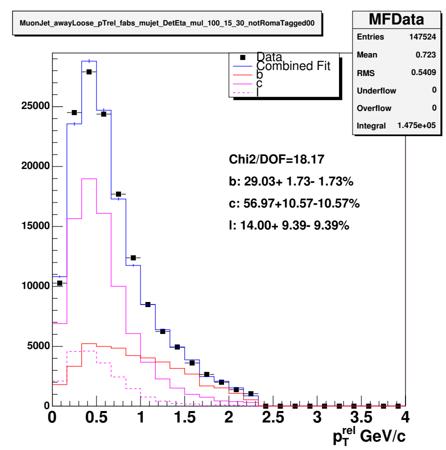


(b) Not Tagged

Figure 106: Fitting b and non-b components to muon p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of $0 \leq \eta < 0.15$.

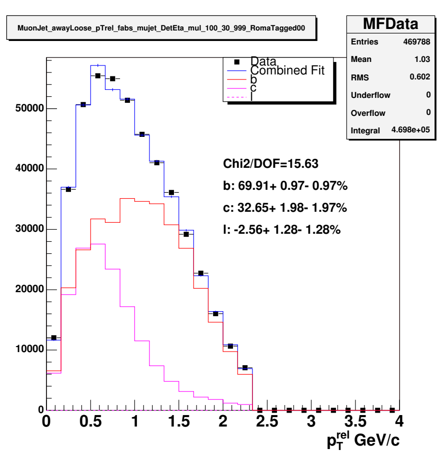


(a) Tagged

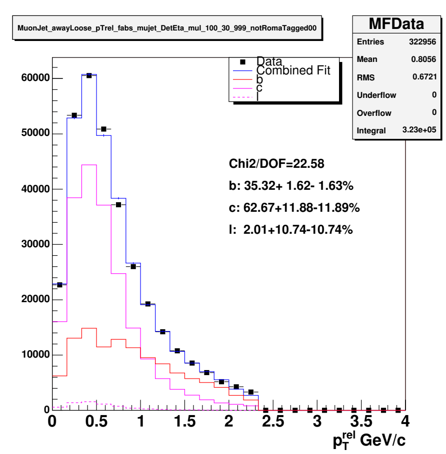


(b) Not Tagged

Figure 107: Fitting b and non-b components to muon p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of $0.15 \leq \eta < 0.30$.

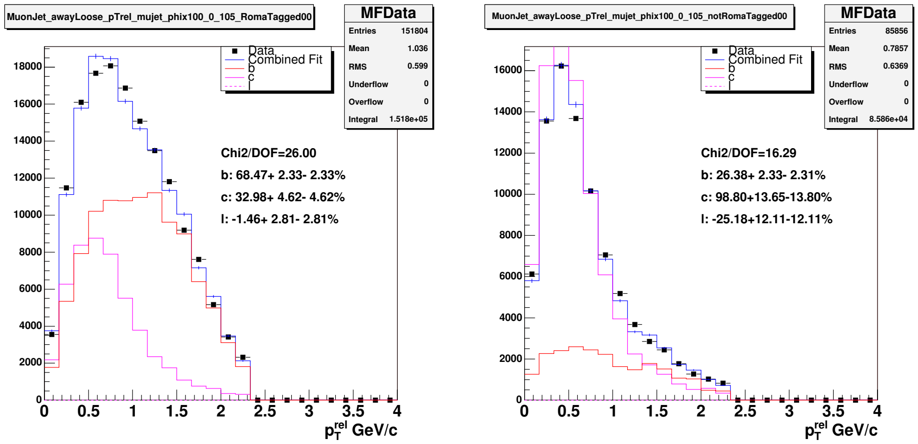


(a) Tagged



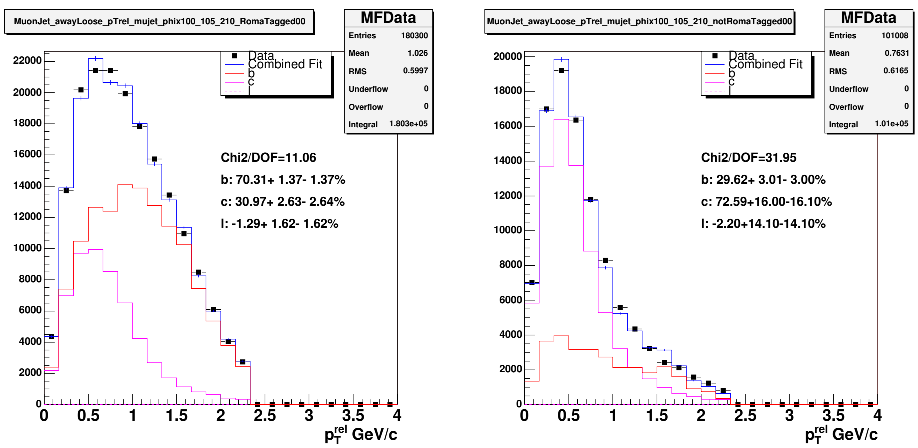
(b) Not Tagged

Figure 108: Fitting b and non-b components to muon p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of $\eta \geq 0.30$.



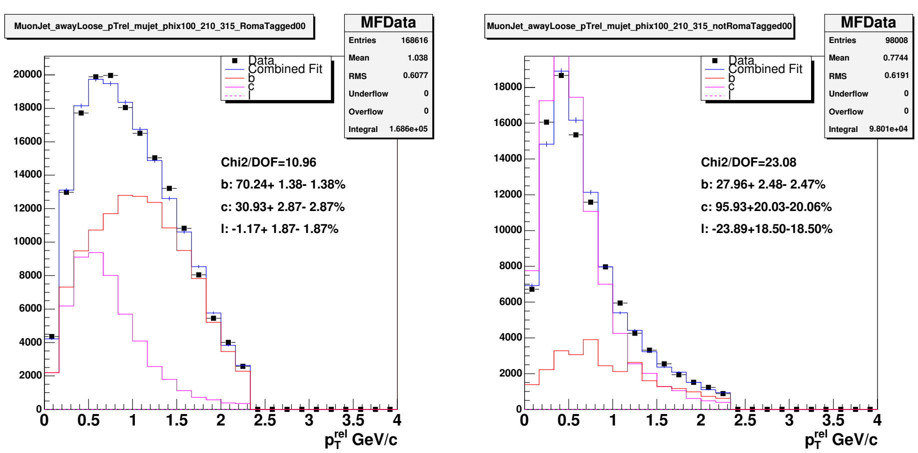
(a) Tagged (b) Not Tagged

Figure 109: Fitting b and non-b components to muon p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of $0 \leq \phi < \frac{\pi}{3}$.



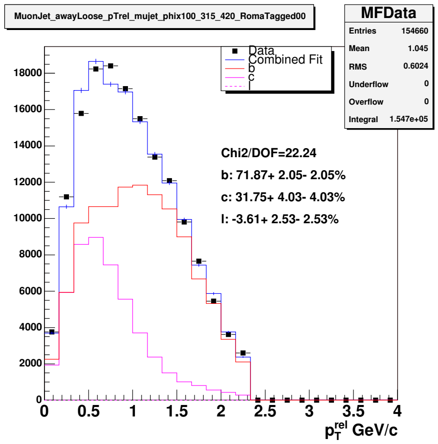
(a) Tagged (b) Not Tagged

Figure 110: Fitting b and non-b components to muon p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of $\frac{\pi}{3} \leq \phi < \frac{2\pi}{3}$.

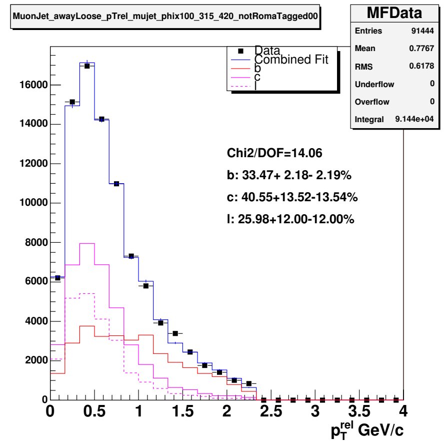


(a) Tagged (b) Not Tagged

Figure 111: Fitting b and non-b components to muon p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of $\frac{2\pi}{3} \leq \phi < \pi$.

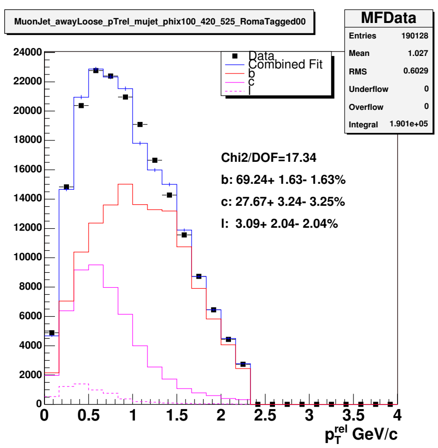


(a) Tagged

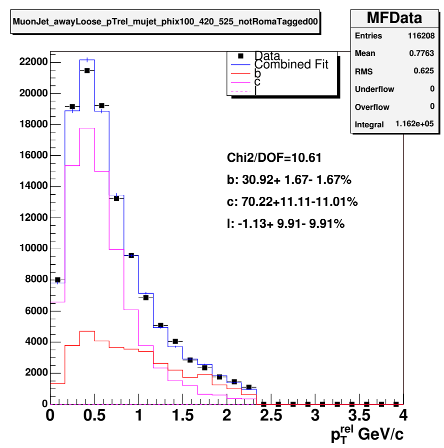


(b) Not Tagged

Figure 112: Fitting b and non-b components to muon p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of $\pi \leq \phi < \frac{4\pi}{3}$.

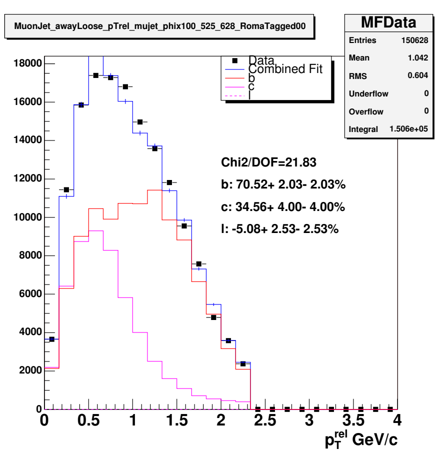


(a) Tagged

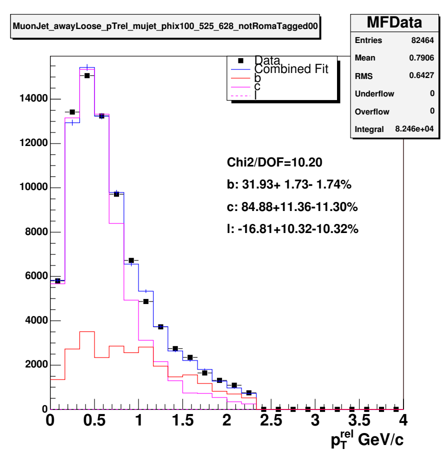


(b) Not Tagged

Figure 113: Fitting b and non-b components to muon p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of $\frac{4\pi}{3} \leq \phi < \frac{5\pi}{3}$.

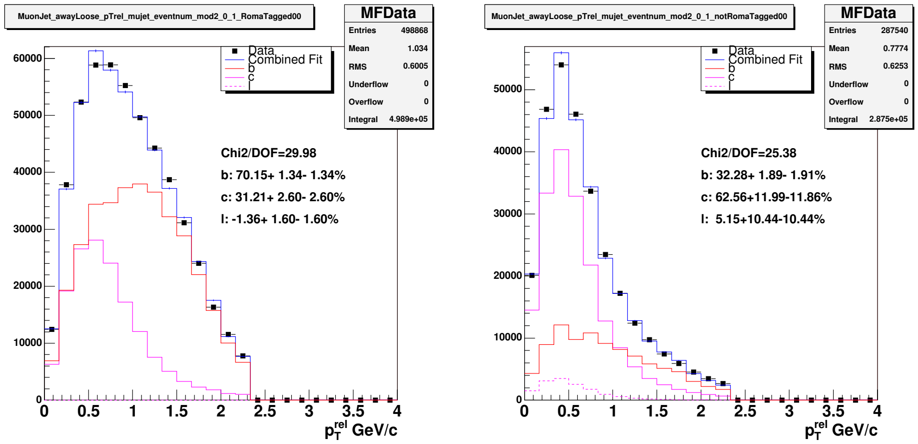


(a) Tagged



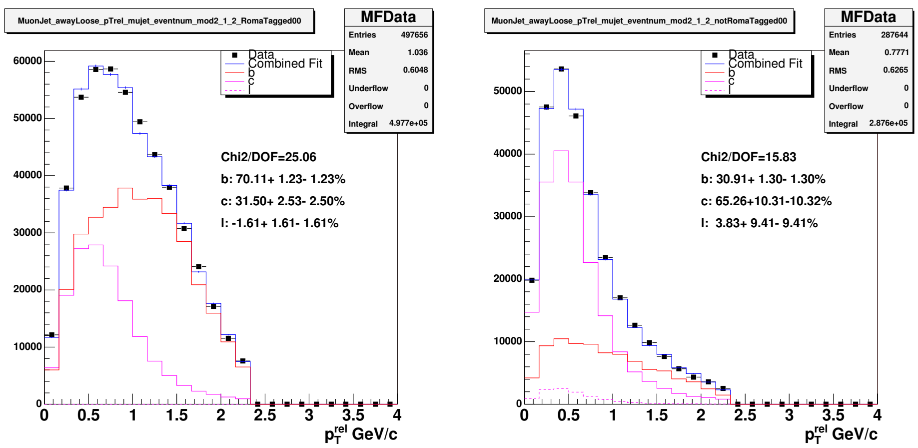
(b) Not Tagged

Figure 114: Fitting b and non-b components to muon p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Bin of $\frac{5\pi}{3} \leq \phi < 2\pi$.



(a) Tagged (b) Not Tagged

Figure 115: Fitting b and non-b components to muon p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Even event number bin.

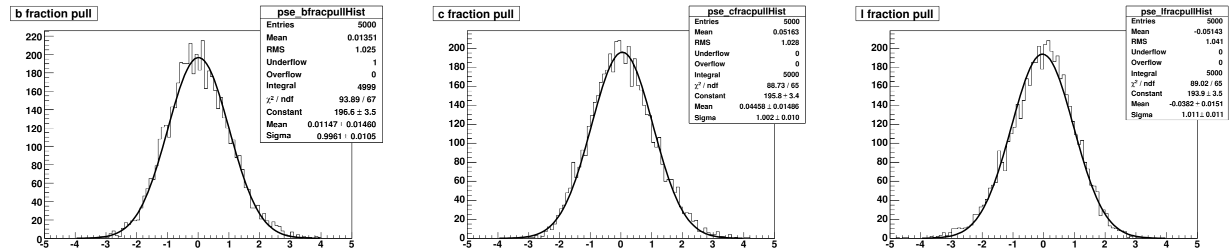


(a) Tagged (b) Not Tagged

Figure 116: Fitting b and non-b components to muon p_T^{rel} data, using definition of RomaNNout3>0.0 for tagged. Odd event number bin.

8.4.3 Pseudoexperiments

Sanity check of fit code.

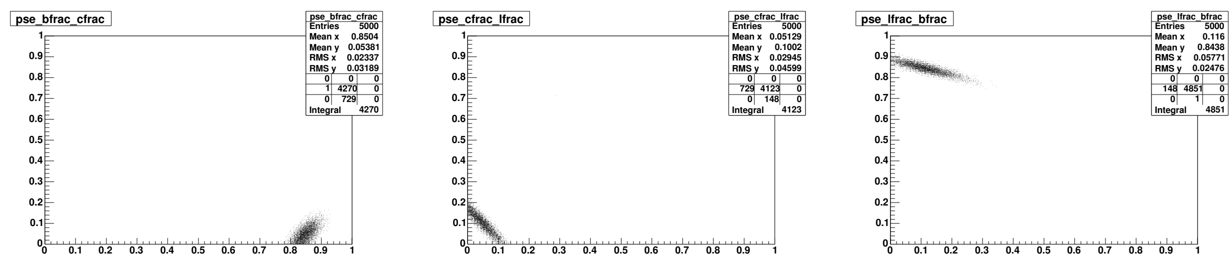


(a) Pseudoexperiments: pull of fitted b-fraction.

(b) Pseudoexperiments: pull of fitted c-fraction.

(c) Pseudoexperiments: pull of fitted l-fraction.

Figure 117: Verification plots using same templates to generate pseudo-data as fitted, showing that the pseudoexperiment framework is valid, that $< f_{b,c,l}^{fitted} > \approx < f_{b,c,l}^{generated} >$ and $\sigma_{pull}^{f_{b,c,l}} \approx 1$. Using default elec p_T^{rel} templates.

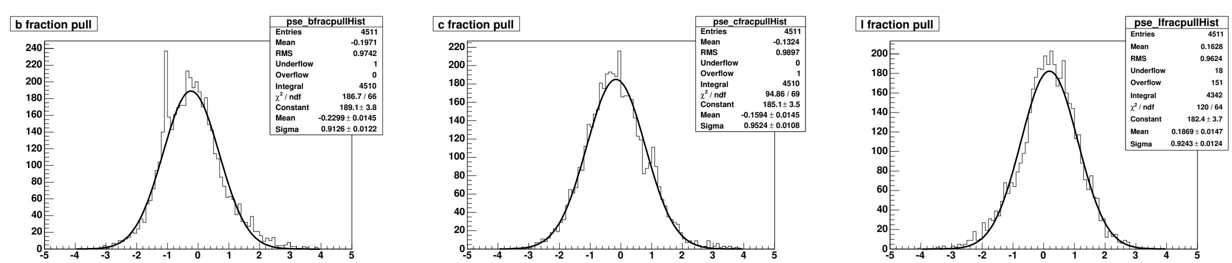


(a) Pseudoexperiments: correlation between b-fraction (x-axis) and c-fraction (y-axis). Correlation factor = 0.61.

(b) Pseudoexperiments: correlation between c-fraction (x-axis) and l-fraction (y-axis). Correlation factor = -0.86.

(c) Pseudoexperiments: correlation between l-fraction (x-axis) and b-fraction (y-axis). Correlation factor = -0.87.

Figure 118: Verification plots using same templates to generate pseudo-data as fitted, showing the correlation between $< f_{b,c,l}^{fitted} >$. Using default elec p_T^{rel} templates.

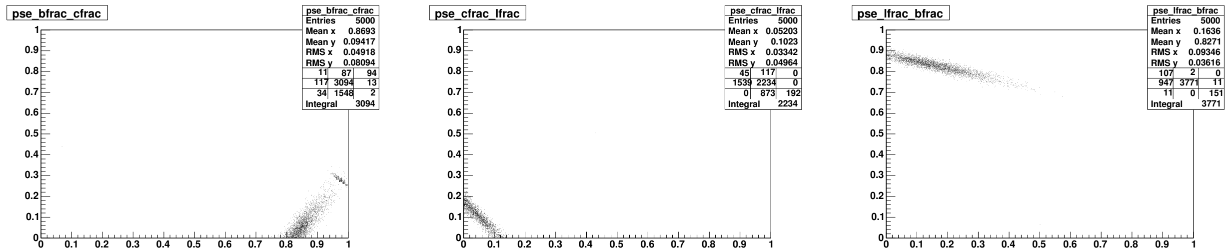


(a) Pseudoexperiments: pull of fitted b-fraction.

(b) Pseudoexperiments: pull of fitted c-fraction.

(c) Pseudoexperiments: pull of fitted l-fraction.

Figure 119: Verification plots using same templates to generate pseudo-data as fitted, showing that the pseudoexperiment framework is valid, that $< f_{b,c,l}^{fitted} > \approx < f_{b,c,l}^{generated} >$ and $\sigma_{pull}^{f_{b,c,l}} \approx 1$. Using elec p_T^{rel}/E_T templates.

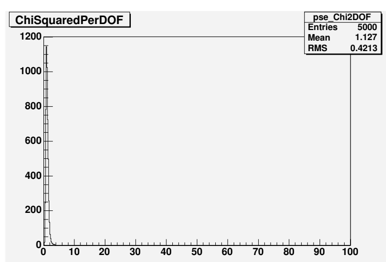


(a) Pseudoexperiments: correlation between b-fraction (x-axis) and c-fraction (y-axis). Correlation factor = 0.84.

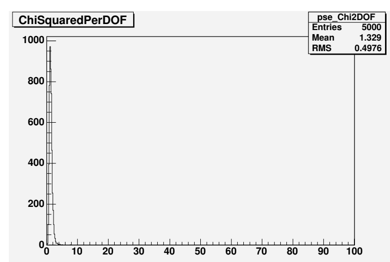
(b) Pseudoexperiments: correlation between c-fraction (x-axis) and l-fraction (y-axis). Correlation factor = -0.84.

(c) Pseudoexperiments: correlation between l-fraction (x-axis) and b-fraction (y-axis). Correlation factor = -0.88.

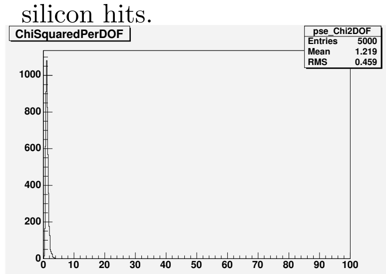
Figure 120: Verification plots using same templates to generate pseudo-data as fitted, showing the correlation between $< f_{b,c,l}^{fitted} >$. Using elec p_T^{rel}/E_T templates.



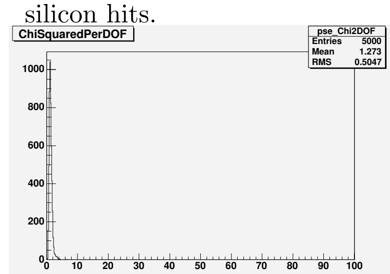
(a) Pseudo-data generated with conversion template of 0 silicon hits.



(b) Pseudo-data generated with conversion template of 3 silicon hits.

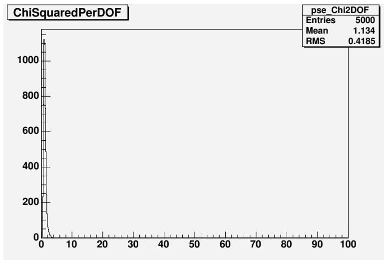


(c) Pseudo-data generated with conversion template of 4 silicon hits.

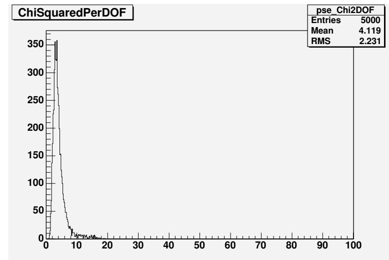


(d) Pseudo-data generated with conversion template of 5 silicon hits.

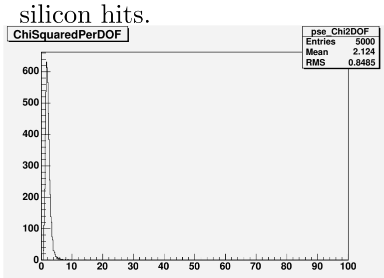
Figure 121: Distribution of χ^2/DOF from elec p_T^{rel} pseudoexperiment fits, with pseudo-data generated from default tagged b-template, default c-template, and 4 different conversion templates. Fitted with default tagged b-template, default c-template, and default conversion template.



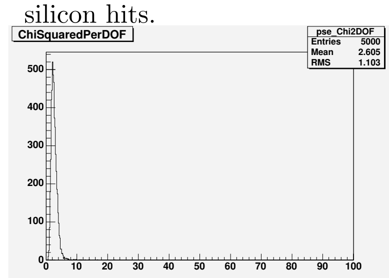
(a) Pseudo-data generated with conversion template of 0 silicon hits.



(b) Pseudo-data generated with conversion template of 3 silicon hits.

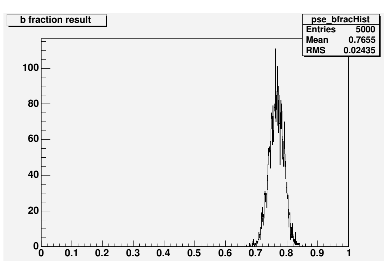


(c) Pseudo-data generated with conversion template of 4 silicon hits.

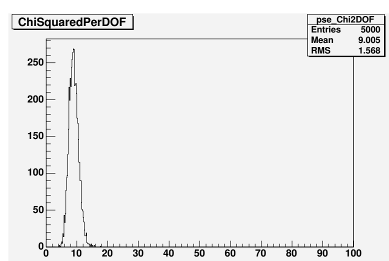


(d) Pseudo-data generated with conversion template of 5 silicon hits.

Figure 122: Distribution of χ^2/DOF from elec p_T^{rel} pseudoexperiment fits, with pseudo-data generated from default not-tagged b-template, default c-template, and 4 different conversion templates. Fitted with default not-tagged b-template, default c-template, and default conversion template.

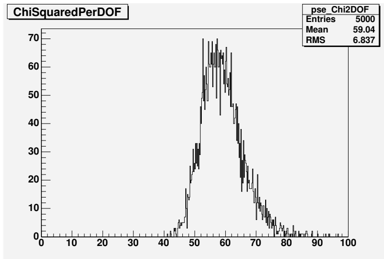


(a) Pseudoexperiments: fitted b-fraction.

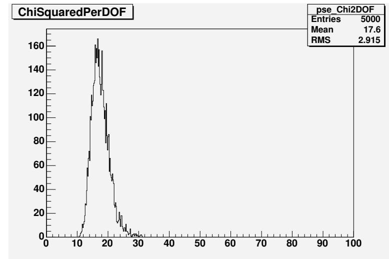


(b) Distribution of χ^2/DOF from pseudoexperiment fit.

Figure 123: Elec p_T^{rel} Pseudoexperiments, with pseudo-data generated with tagged data-based b-template, default c-template, and default conversion templates. Fitted with tagged MC-based b-template, default c-template, and default conversion templates.

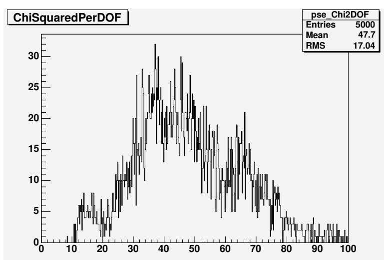


(a) Pseudo-data generated with tagged Peterson (PFP=0.0025) b-template.

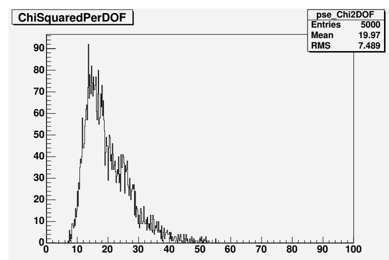


(b) Pseudo-data generated with tagged Peterson (PFP=0.0041) b-template.

Figure 124: Distribution of χ^2/DOF from elec p_T^{rel} pseudoexperiment fits, with pseudo-data generated with tagged b-templates from $W + b\bar{b}$ MC with 2 different Peterson fragmentation modelling, default c-template, and default conversion templates. Fitted with tagged b-template from $W + b\bar{b}$ MC with Bowler-Lund fragmentation modelling, default c-template, and default conversion template.

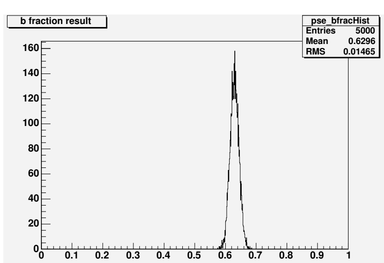


(a) Pseudo-data generated with not-tagged Peterson (PFP=0.0025) b-template.

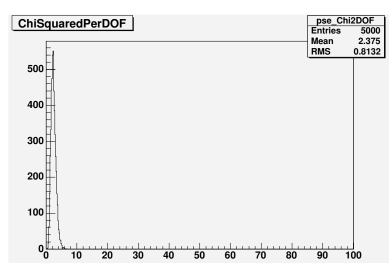


(b) Pseudo-data generated with not-tagged Peterson (PFP=0.0041) b-template.

Figure 125: Distribution of χ^2/DOF from elec p_T^{rel} pseudoexperiment fits, with pseudo-data generated with not-tagged b-templates from $W + b\bar{b}$ MC with 2 different Peterson fragmentation modelling, default c-template, and default conversion templates. Fitted with not-tagged b-template from $W + b\bar{b}$ MC with Bowler-Lund fragmentation modelling, default c-template, and default conversion template.

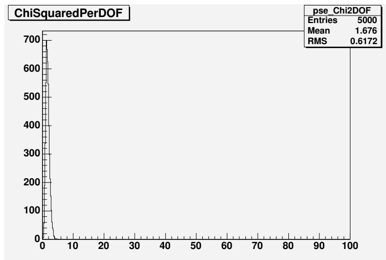


(a) Pseudoexperiments: fitted b-fraction.

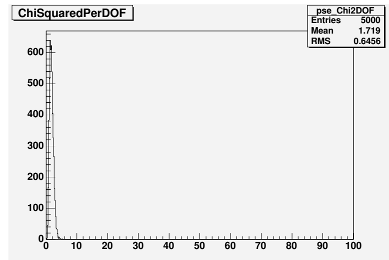


(b) Distribution of χ^2/DOF from pseudoexperiment fit.

Figure 126: Muon p_T^{rel} pseudoexperiments, with pseudo-data generated with tagged data-based b-template, default c-template, and default conversion templates. Fitted with tagged MC-based b-template, default c-template, and default conversion templates.

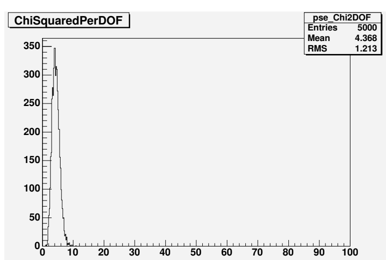


(a) Pseudo-data generated with tagged Peterson (PFP=0.0025) b-template.

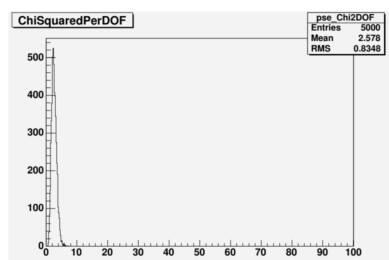


(b) Pseudo-data generated with tagged Peterson (PFP=0.0041) b-template.

Figure 127: Distribution of χ^2/DOF from muon p_T^{rel} pseudoexperiment fits, with pseudo-data generated with tagged b-templates from $W + b\bar{b}$ MC with 2 different Peterson fragmentation modelling, default c-template, and default conversion templates. Fitted with tagged b-template from $W + b\bar{b}$ MC with Bowler-Lund fragmentation modelling, default c-template, and default conversion template.



(a) Pseudo-data generated with not-tagged Peterson (PFP=0.0025) b-template.



(b) Pseudo-data generated with not-tagged Peterson (PFP=0.0041) b-template.

Figure 128: Distribution of χ^2/DOF from muon p_T^{rel} pseudoexperiment fits, with pseudo-data generated with not-tagged b-templates from $W + b\bar{b}$ MC with 2 different Peterson fragmentation modelling, default c-template, and default conversion templates. Fitted with not-tagged b-template from $W + b\bar{b}$ MC with Bowler-Lund fragmentation modelling, default c-template, and default conversion template.

References

[1] Physics Information Technologies GmbH, http://www.phi-t.de/

[2] F. Garberson *et al.* "SecVtx B-Tag Efficiency Measurement Using Muon p_T^{rel} for 1.2 fb^{-1} Analyses", CDF Note 8640.

[3] T. Wright *et al.* "SecVtx B-Tag Efficiency Using Muon Transverse Momentum", CDF Note 7448.

[4] D. Jeans *et al.* "A new b & c jet identification algorithm", CDF Note 8451.

[5] Sandro De Cecco "High PT Heavy Flavor Tagging", CDF Collaboration Meeting 2007.

[6] P. Mastrandrea *et al.* "Calibration Procedure for the Roma Neural Network Heavy Flavor Tagging Algorithm", CDF Note 8764.

[7] Lucio Cerrito, Anyes Taffard "A Soft Muon Tagger for Run II", CDF Note 7122.

[8] F. Garberson *et al.* "Combination of the SecVtx 1.2 fb^{-1} b-Tagging Scale Factors", CDF Note 8666.

[9] B.Alvarez *et al.* "Jet Probability Tag Rate Matrix for Gen6 2.2invfb Data", CDF Note 9227.

[10] JP Chou, Private Communications, 2008.

[11] Jean-Francois Arguin, Pekka K. Sinervo "b-jet Energy Scale Uncertainty From Existing Experimental Constraints", CDF Note 7252.